OM nucleic - nucleic search, using sw model

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:          June 15, 2002, 09:52:04 ; Search time 1902.97 Seconds
                                            (without alignments)
                                            1971.736 Million cell updates/sec

Title:           US-09-445-362B-6
Perfect score:   278
Sequence:        1 gccaacacgcantccgacga..........cattttnaactccccaggacc 278

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters:      27472414

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :      EST:*
                1: em_estba:*
                2: em_esthum:*
                3: em_estin:*
                4: em_estmu:*
                5: em_estov:*
                6: em_estpl:*
                7: em_estro:*
                8: em_htc:*
                9: gb_est1:*
               10: gb_est2:*
               11: gb_htc:*
               12: gb_gss:*
               13: em_gss_hum:*
               14: em_gss_inv:*
               15: em_gss_pln:*
               16: em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

C	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	257.8	92.7	279	10	C04498	C04498 Huma
	2	214.4	77.1	477	9	AI327454	AI327454 ma68a06.y
	3	203	73.0	416	10	BF543291	BF543291 UI-R-Y0-a
	4	193	69.4	468	10	W12756	W12756 ma68a06.r1
c	5	150.4	54.1	459	12	AZ985110	AZ985110 2M026602
c	6	150.4	54.1	593	9	AA624918	AA624918 vn90e12.r
	7	149	53.6	608	10	BF454139	BF454139 maa10c03.
	8	145.6	52.4	616	9	AA881875	AA881875 vx29a04.r
c	9	145	52.2	309	10	C04505	C04505 Huma
	10	141	50.7	938	10	BI853762	BI853762 603380312
	11	137.2	49.4	538	10	BF140778	BF140778 601788336
c	12	131.2	47.2	518	12	AQ311743	AQ311743 RPCII1-10
c	13	120	43.2	638	10	BJ068008	BJ068008 UI-M-CG0p
	14	115.6	41.6	960	9	AV254558	AV254558 AV254558
	15	115.6	41.6	467	10	AK017072	AK017072 Mus muscu
	16	115.6	41.6	1635	11	AK017072	AK017072 PM2-CN009
	17	112.4	40.4	337	10	BG979894	BG979894
	18	107.8	38.8	978	10	BE256701	BE256701 601115503
c	19	104.8	37.6	520	9	AI429173	AI429173 m154a06.x
	20	104	37.4	582	9	AA031252	AA031252 m129f07.r
	21	103.6	37.3	1600	11	AK015263	AK015263 Mus muscu
	22	103.4	37.2	505	10	W61419	W61419 md92f06.r1
	23	103.4	37.2	533	10	AA626666	AA626666 ab51e06.r
	24	101.8	36.6	684	10	BG700839	BG700839 602681741
	25	101.8	36.6	754	9	AI885163	AI885163 w190e03.x
	26	101.4	36.5	504	9	AA455871	AA455871 aa01c12.r
	27	100.2	36.0	757	10	BG703244	BG703244 602685323
c	28	99.2	35.7	660	9	AA885078	AA885078 am30g04.s
	29	99.6	35.5	510	10	BF513815	BF513815 UI-H-BW1-
c	30	94.4	34.0	249	9	AA936258	AA936258 on75b07.s
	31	94.4	34.0	490	10	BF335156	BF335156 RC0-CT042
c	32	94.4	34.0	525	9	AA203464	AA203464 zx59b11.r
	33	94.4	34.0	588	9	AA143016	AA143016 z148h02.r
	34	94.4	34.0	751	10	BF791990	BF791990 602252383
	35	94.2	33.9	865	10	BG529209	BG529209 602558283
	36	94	33.8	513	9	AI145620	AI145620 UI-R-B10-
	37	94	33.8	784	9	AA877607	AA877607 nt07g10.s
c	38	93.8	33.7	872	9	AL562582	AL562582 AL562582
c	39	93.8	33.7	710	9	AI022657	AI022657 ox05h04.x
c	40	92.6	33.3	622	10	BF347015	BF347015 60202803
c	41	90.2	32.4	448	10	W54078	W54078 md12a08.r1
	42	89.6	32.2	836	10	BG673258	BG673258 DRNBRB11
	43	89.4	32.2	435	9	AA636169	AA636169 vn15c12.r
c	44	87.8	31.6	664	9	AI322547	AI322547 mi29f07.y
	45	86.8	31.2	480	10	BF557530	BF557530 UI-R-E1-g

ALIGNMENTS

```
RESULT  1
C04498

LOCUS       C04498                  279 bp    mRNA    linear   EST 30-JUL-1996
DEFINITION  C04498 Human heart cDNA (YNakamura) Homo sapiens cDNA clone
            3NHC3467, mRNA sequence.
ACCESSION   C04498
VERSION     C04498.1  GI:1467749
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
AUTHORS     Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and
            Nakamura,Y.
TITLE       Construction of a normalized directionally cloned cDNA library fr
            adult heart and analysis of 3040 clones by partial sequencing
JOURNAL     Genomics 35 (1), 231-235 (1996)
MEDLINE     96299762
COMMENT     Contact: Yusuke Nakamura
            Institute of Medical Science
            University of Tokyo
            4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
            Tel: 81-3-5449-5372
            Fax: 81-3-5449-5433
            Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1..279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="3NHC3467"
                     /clone_lib="Human heart cDNA (YNakamura)"
                     /dev_stage="adult"
                     /note="Organ: heart; normalized directionally cloned cDNA
                     from adult heart"
BASE COUNT       79 a     75 c     71 g     48 t     6 others
ORIGIN

Query Match    92.7%;  Score 257.8;  DB 10;  Length 279;
```

RESULT 2

Best Local Similarity 98.6%;   Pred. No. 7e-54;   DB 3;
Matches 275;   Conservative 0;   Mismatches 3;   Indels 1;   Gaps 1;

```
Qy    1 gccaacacgcantccgacgacagtgcagccaatggtcattgcagagatgc-ncaaagtcaa  59
         ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db    1 GCCAACACGCANTCCGACGACAGTGCAGCCAATGGTCATTGCAGAGATGCNTCAAAGTCAA  60

Qy   60 tgagcaacatccacacagtaaacgtcgagtcccacaacttcatcatacgggaaagggtggc 119
         ||||||||||||| |||||| |||||||||||||| ||||||||| ||| ||||||||||
Db   61 TGAGCAACATCCACCACAGTCAAACGTCGAGTCCCAACTTCATCAACGGGGAAAGGGTGGC 120

Qy  120 catcatgagagctctccagcaacaacaacggtgctcacggagctcggtgtttcataaccagag 179
         |||||||||||||||||||||||||||| ||||| ||||||||||| |||||||||||||
Db  121 CATCATGAGAGCTCTCCAGCAACAACACGGTGCTCCACGGAGCTCGGTTTCATAACCAGAG 180

Qy  180 gcaccatcatgggcagccaggtggaaatggagattgtcaagctnctgaaggagaacacgac 239
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GCAACATCATGGGCAGCCAGGTGGAAATGGAGATTGTCAAGCTNCTGAAGGAGAACACGAC 240

Qy  240 gctnctgaggctggntaccattttnaactcccaggacc 278
         |||||||||||||| ||||||||||||||||||||||
Db  241 GCTNCTGAGGCTGGGNTACCATTTTNAACTCCCAGGACC 279
```

LOCUS   AI327454   477 bp   mRNA   linear   EST 23-DEC-1998
DEFINITION   ma68a06.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
             IMAGE:315826 5' similar to gb:X54162_rna1 64 KD AUTOANTIGEN D1
             (HUMAN);, mRNA sequence.
ACCESSION   AI327454
VERSION   AI327454.1 GI:4061883
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 477)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:206442
             This read is a RESEQUENCE of a previously sequenced mouse clone.
             This read has been verified (found to hit its original self in the
             correct orientation).
             Seq primer: -40RP from Gibco
             High quality sequence stop: 459.
FEATURES   Location/Qualifiers
     source   1..477
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:315826"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAAGACCGCGCATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."

BASE COUNT   147 a   119 c   125 g   86 t
ORIGIN

Query Match
Best Local Similarity 77.1%;   Score 214.4;   DB 9;   Length 477;
         84.9%;   Pred. No. 4.6e-43;
Matches 236;   Conservative 0;   Mismatches 42;   Indels 0;   Gaps 0;

```
Qy    1 gccaacacgcantccgacgacagtgcagccaatggtcattgcagagatgc-ncaaagtcaat 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db    1 ...                                                            60

Qy   61 gagcaacatccacaacgtaaacgtcgagtcccaacttcatcatacgggaaagggtggc 120
Db   99 GCCAACACGCGCGCGCGGGATTGCACATTGCAGAGACATGCTCAAGGTTAAC 158

Qy  121 atcatgagagctctccagcaacaacaacggtgctcacggagctcggtgtttcataaccagag 180
Db  159 GAGCACATTACCACGGTAAACGTCGAGTCCACTCATCACTGGCCAACAACCAGAG 218

Qy  181 gcaccatcatgggcagccaggtggaaatggagattgtcaagctnctgaaggagaacacgac 240
Db  279 CACATCATGGGCGCTCCAGGTGGAGAATTGCGAAATTGTCAAGCTGCTCAAAGAGAATACCACG 338

Qy  241 ctnctgaggctggntaccattttnaactcccaggacc 278
         |||||||||||||| |||||||| |||||||||||||
Db  339 CTGCTGAGGCTGGGCTACCATTTGAGCTCCCAGGACC 376
```

RESULT 3

Best Local Similarity 84.9%;   Pred. No. 4.6e-43;   DB 9;
Matches 236;   Conservative 0;   Mismatches 42;   Indels 0;   Gaps 0;

LOCUS   BF543291   416 bp   mRNA   linear   EST 11-DEC-2000
DEFINITION   UI-R-Y0-abd-c-03-0-UI.r1 UI-R-Y0 Rattus norvegicus cDNA clone
             UI-R-Y0-abd-c-03-0-UI 5', mRNA sequence.
ACCESSION   BF543291
VERSION   BF543291.1 GI:11634398
KEYWORDS   EST.
SOURCE   Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 416)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE   Normalization and subtraction: two approaches to facilitate gene
             discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT   Contact: Soares, MB
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             cDNA Library Preparation: M.B. Soares Lab Clone distribution:
             clones will be available through Research Genetics (www.resgen.com)
             This clone is also available through the I.M.A.G.E. Consortium at
             LLNL (info@image.llnl.gov). IMAGE ID= 1790895
             Seq primer: M13 Forward.
FEATURES   Location/Qualifiers
     source   1..416
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-Y0-abd-c-03-0-UI"
                     /clone_lib="UI-R-Y0"
```

/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."

BASE COUNT    123 a   112 c   105 g   76 t
ORIGIN

Query Match

Best Local Similarity   84.8%;    Score 203; DB 10;    Length 416;
Matches 224; Conservative   0;   Mismatches   40;   Indels   0;   Gaps   0;
    73.0%;     Pred. No. 2.9e-40;

```
Qy    1 gccaacagcgcantccgacgacaagtgcagccatggtcattgcagagatgcncaaagtcaat  60
         ||||||||||| ||||||  ||||||||||||||||||||||||||||||  ||||||||||
Db  152 GCCAACACGCACGCTGAGCGACGCCGCGGATTGCCATTGCGGAAATGCTCAAGGTTAAC 211

Qy   61 gagcacatcaccaaacgtgaaacgtcgagtccaacttcatacgggaaagggatcctggcc  120
         ||||||  |||| |||| || |||||||||||||||||| || ||||||  ||||||||||
Db  212 GAGCACAATTACCACGTAAACGTGGAGTCCAACTTCATCACCGGGAAGGGGATCCTGGCC 271

Qy  121 atcatgagagctctccgcacacaacacggtgtctcacggagctgcgggttcacaaccagg  180
         |||||||||||  | ||||| |||| || || ||||||||||||| ||||| |||||| |
Db  272 ATCATGAGAGCGTGCACGACACACAGTGCTCACCGAGCTGCGCGTTCCACAACCAGAGG 331

Qy  181 cacatcatgggcagccaggtgggaaatggagattgtcaagctnctgaaggagaacacgacg  240
         ||||| |||||| |  | |  |||| |||||| ||||||| ||  ||| ||||  |||||||
Db  332 CACATCAATGGGCGTCCCGTGCGACCAACACGGTGCGTCCAAGGGGGATTCCTGGCC 391

Qy  241 ctnctgaggctgggntaccatttg  264
         ||  |||||||||| |||||||||
Db  392 CTGCTGAGGCTGGGCTACCATTTT 415
```

RESULT   4
LOCUS   W12756
W12756
DEFINITION   ma68a06.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:315826 5' similar to gb:X54162_rna1 64 KD AUTOANTIGEN D1
(HUMAN);, mRNA sequence.
ACCESSION   W12756
VERSION   W12756.1   GI:1286794
KEYWORDS   EST.
SOURCE   house mouse.
   ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1   (bases 1 to 468)
   AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
     Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
     Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
     Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
     Waterston,R.
   TITLE   The WashU-HHMI Mouse EST Project

LOCUS   W12756   468 bp   mRNA   linear   EST 02-OCT-1997

JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
     WashU-HHMI Mouse EST Project
     Washington University School of Medicine
     4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
     Tel: 314 286 1800
     Fax: 314 286 1810
     Email: mouseest@watson.wustl.edu
     This clone is available royalty-free through LLNL ; contact the
     IMAGE Consortium (info@image.llnl.gov) for further information.
     MGI: 206442
     Seq primer: mob.REGA+ET
     High quality sequence stop: 403.

FEATURES   Location/Qualifiers
   source   1..468
         /organism="Mus musculus"
         /db_xref="taxon:10090"
         /clone="IMAGE:315826"
         /clone_lib="Soares mouse p3NMF19.5"
         /dev_stage="19.5 dpc total fetus"
         /lab_host="DH10B (ampicillin resistant)"
         /note="Vector: pT7T3D (Pharmacia) with a modified
         polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
         was primed with a Not I - oligo(dT) primer [5'
         TGTTACCAATCTGAAGTGGGAGCGCGCCGAATTTTTTTTTTTTTTTTT 3'],
         adapters (Pharmacia), digested with Not I and cloned into
         the Not I and Eco RI sites of a modified pT7T3 vector
         (Pharmacia). Library went through one round of
         normalization to a Cot = 5. Library constructed by Bento
         Soares and M.Fatima Bonaldo. RNA was kindly provided by
         Dr. Minoru Ko (Wayne State University)."

BASE COUNT   144 a   118 c   120 g   85 t   1 others
ORIGIN

Query Match

Best Local Similarity   84.6%;   Score 193; DB 10;   Length 468;
Matches 236; Conservative   0;   Mismatches   41;   Indels   2;   Gaps   2;
    69.4%;   Pred. No. 8.9e-38;

```
Qy    1 gccaacagcgcantccgacgacaagtgcagccatggtcattgcagagatgcncaaagtcaat  60
         ||||||||||| |||||| ||||||||||||||||||||||||||||||| ||||||||||
Db   97 GCCAACACGCACGCTGATTGACGCCGC-GCGATTGCCATTGCAGACAATGCTCAAGGTTAAC 155

Qy   61 gagcacatcaccaaacgtgaaacgtcgagtccaacttcatacgggaaagggatcctggcc  120
         ||||||  |||| |||| || ||||||||||||||||||| || ||||||  ||||||||||
Db  156 GAGCACAATTACCACGTAAACGTGGAGTCCAACTTCATCACCGGCAAGGGGATCCTGGCC 215

Qy  121 atcatgagagctctccgcacacaacacggtgtctcacggagctgcgggttcacaaccagg  180
         |||||||||||  | ||||| |||| || || ||||||||||||| ||||| |||||| |
Db  216 ATCATGAGAGCCCTGCACGACACACGGTGCGTCACGGAGCTGCGCTTCCACAACCAGCGG 275

Qy  181 cacatcatgggcagccaggtgggaaatggagattgtcaagctnctgaaggagaacacgacg  240
         ||||| |||||| |  | |  |||| |||||| ||||||| ||  ||| ||||  |||||||
Db  276 CACATCAATGGGCTTCGCAGCTGGAGATGGAGAAATTGTCAAGCTGCTCAAGGAGAGAATACCACG 335

Qy  241 ctnctgaggctgggntaccatt-ttnaactcccaggacc  278
         ||  |||||||||| |||||||  ||  ||||||| |||
Db  336 CTGCTGAGGCTGGGCTACCATTNTTGAGCTCCAGGAC 374
```

RESULT   5
LOCUS   AZ985110/c
AZ985110
DEFINITION   2M02660C19R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M02660C19 R, DNA sequence.
ACCESSION   AZ985110
VERSION   AZ985110.1   GI:11856337
KEYWORDS   GSS.
SOURCE   house mouse.
   ORGANISM   Mus musculus
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

LOCUS   AZ985110   459 bp   DNA   linear   GSS 27-APR-2001

REFERENCE   1 (bases 1 to 459)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacom,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D., Weiss,R.
TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000   Std Error: 0.00
             Insert length: 10000
             Plate: 0266   row: C   column: 19
             Seq primer: CACACAGGAAACAGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 459.
FEATURES   Location/Qualifiers
             source   1..459
                         /organism="Mus musculus"
                         /strain="C57BL/6J"
                         /db_xref="taxon:10090"
                         /clone="UUGC2M0266C19"
                         /clone_lib="Mouse 10kb plasmid UUGC2M library"
                         /sex="Female"
                         /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                         /note="Vector: PWD42nv; Purified genomic DNA from M.
                         musculus C57BL/6J (female) was obtained from the Jackson
                         Laboratory Mouse DNA Resource
                         (http://www.jax.org/resources/documents/dnares/). The DNA
                         was hydrodynamically sheared by repeated passage through a
                         0.005 inch orifice at constant velocity. The sheared DNA
                         was blunt end-repaired with T4 DNA polymerase and T4
                         polynucleotide kinase. Adaptor oligonucleotides were
                         ligated to the blunt ends in high molar excess. The
                         adaptored DNA was purified and size-selected for a 9.5 to
                         10.5 kb range using preparative agarose gel
                         electrophoresis. Vector DNA was prepared from a derivative
                         of pWD42 (gi|14732114|gb|AF129072.1|, a copy-number
                         inducible derivative of plasmid R1. The vector was ligated
                         with adaptors complementary to the insert adaptors and
                         purified. The sheared, adaptored mouse DNA was annealed to
                         adaptored vector DNA, and transformed into
                         chemically-competent E. coli XL10-Gold (Stratagene) cells
                         and selected for ampicillin resistance."

BASE COUNT   83 a   121 c   132 g   123 t
ORIGIN

Query Match   54.1%; Score 150.4; DB 12; Length 459;
Best Local Similarity 70.5%; Pred. No. 3e-27;
Matches 196; Conservative   0; Mismatches 82; Indels   0; Gaps   0;

Qy   1 gccaacacgcantccgacgacagtgcagccatgtcattgcagagatgcncaaagtcaat 60
           ||||||||| ||||| |||||||||||||||||||||||||||||||| |||||||||||
Db   398 GCCAATACTCGGGCTGACGACCATGTGGCCCTTTGCAATTGCTATCATGTTAAGGCCAAT 339

Qy   61 gagcacatcaccaacgtaaacgtcgagtccaacttcatacggaaaggggatcctggcc 120
           || |||||||||||||| |||||||||||||| ||||||||||||||||||||||||||
Db   338 AAGACCATCACCAGGCTCACCTGGACTCCAATCACTGCCAAAGGCATCCTGCC 279

Qy   121 atcatgagagctccagcacacaacggtgctcacggagctgcggttcataaccagagg 180
           ||||||||||| ||||||||| ||||||||||| ||||||||||||||||||| |||||
Db   278 ATCTTCCGGCGCGCCTCCTCAGACAACACTCTGACAGAGCTGCGGCTTCACCAGAGG 219

Qy   181 cacatcatgggcacggcaggtggaaaatggagggaagaacaacacg 240
           ||||| ||||||||||||||| ||||||||||||||||||||||| |
Db   218 CACATCTGTGGGGCAGGGCAAGAGTCGCCAAGACTGAGATGGAGAATCGGCAAGGAGAGAACAACACC 159

RESULT   6
LOCUS   AA624918   593 bp   mRNA   linear   EST 14-OCT-1997
DEFINITION   vn90e12.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
             IMAGE:1039246 5' similar to gb:X54162_rna1 64 KD AUTOANTIGEN D1
             (HUMAN);, mRNA sequence.
ACCESSION   AA624918
VERSION   AA624918.1   GI:2528794
KEYWORDS   EST.
SOURCE   Mus musculus
  ORGANISM   Mus musculus
             house mouse.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 593)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
             Washu-HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mousesest@watson.wustl.edu
             This clone is available royalty-free through LLNL.; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:579270
             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 446.
FEATURES   Location/Qualifiers
             source   1..593
                         /organism="Mus musculus"
                         /strain="NIH/Swiss"
                         /db_xref="taxon:10090"
                         /clone="IMAGE:1039246"
                         /clone_lib="Stratagene mouse heart (#937316)"
                         /sex="pooled"
                         /tissue_type="heart"
                         /dev_stage="13 day embryos"
                         /lab_host="SOLR (kanamycin resistant)"
                         /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                         EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                         Oligo dT: 93 pooled NIH/Swiss 13 day embryo hearts.
                         Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                         adaptor sequence: 5' GAATTCGGCACGAG 3'-3' adaptor
                         sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'."

BASE COUNT   170 a   169 c   156 g   98 t
ORIGIN

Query Match   54.1%; Score 150.4; DB 9; Length 593;
Best Local Similarity 70.5%; Pred. No. 3.3e-27;
Matches 196; Conservative   0; Mismatches 82; Indels   0; Gaps   0;

Qy   1 gccaacacgcantccgacgacagtgcagccatgtcattgcagagatgcncaaagtcaat 60
           ||||||||| ||||| |||||||||||||||||||||||||||||||| |||||||||||
Db   146 GCCAATACTCGGGCTGACGACCATGTGGCCCTTTGCAATTGCTATCATGTTAAGGCCAAT 205

Qy   61 gagcacatcaccaacgtaaacgtcgagtccaacttcatacggaaaggggatcctggcc 120
           || |||||||||||||| |||||||||||||| ||||||||||||||||||||||||||
Db   206 AAGACCATCACCAGGCTCACCTGGACTCCAATCACTGCCAAAGGCATCCTGCC 265

Qy   121 atcatgagagctccagcacacaacggtgctcacggagctgcggttcataaccagagg 180
           ||||||||||| ||||||||| ||||||||||| ||||||||||||||||||| |||||
Db   120

Qy   241 ctncctgaggctgggntaccattttnaactcccaggacc 278
           | ||| ||||||||| ||||||||||||| |||| |||||
Db   158 TTGCTCAAGCTCGGGCTACCCATTTTGAGCTCGGCCAGGACC 121

```
OM nucleic - nucleic search, using sw model
```

```
Run on:        June 15, 2002, 21:27:21 ; Search time 289.04 Seconds
                                          (without alignments)
                                          1651.336 Million cell updates/sec

Title:         US-09-445-362B-6
Perfect score: 278
Sequence:      1 gccaacacgcantccgacga.........cattttnaactcccaaggacc 278

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched:      1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:   3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
                 Maximum Match 100%

Database :     N_Geneseq_032802:*
               Listing first 45 summaries

    1:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1980.DAT:*
    2:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1981.DAT:*
    3:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1982.DAT:*
    4:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1983.DAT:*
    5:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1984.DAT:*
    6:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1985.DAT:*
    7:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1986.DAT:*
    8:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1987.DAT:*
    9:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1988.DAT:*
   10:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1989.DAT:*
   11:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1990.DAT:*
   12:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1991.DAT:*
   13:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1992.DAT:*
   14:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1993.DAT:*
   15:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1994.DAT:*
   16:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1995.DAT:*
   17:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1996.DAT:*
   18:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1997.DAT:*
   19:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1998.DAT:*
   20:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA1999.DAT:*
   21:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA2000.DAT:*
   22:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA2000A.DAT:*
   22:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA2001A.DAT:*
   23:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA2001B.DAT:*
   24:  /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	265.6	95.5	906	21	AAC76864	Human ORFX ORF2419
2	264.6	95.2	2281	22	AAS31144	Human diagnostic a
3	264	95.0	1936	20	AAV74263	Human heart muscle
4	264	95.0	2080	20	AAV74264	Human heart muscle
5	264	95.0	2268	20	AAV74265	Human heart muscle
6	261	93.9	279	20	AAV74267	Human heart muscle
7	212.6	76.5	274	22	ABA31593	Probe #10059 for g
8	212.6	76.5	274	22	AAK12914	Human brain expres
9	203.6	73.2	489	22	ABA21609	Probe #75 for gene

No.	Score	Match %	Length	DB	ID	Description
10	203.6	73.2	489	22	AAK00088	Human brain expres
11	152	54.7	990	21	AAA50238	DNA encoding autoa
12	152	54.7	1076	21	AAA50240	DNA encoding autoa
13	152	54.7	3913	22	AAI57784	Human polynucleoti
14	152	54.7	4071	22	AAI57998	Human polynucleoti
15	106.2	38.2	524	22	AAH93735	Human protein enco
16	104	37.4	1350	22	AAK52334	Human polynucleoti
17	100.8	36.3	676	22	AAK53318	Human polynucleoti
18	97.6	35.1	1528	22	AAH99177	Human ORFX ORF2864
19	97.6	35.1	1528	22	AAH98259	Human EST-derived
20	94.4	34.0	2124	22	AAI58689	Human EST-derived
21	94.4	34.0	2503	21	AAI60475	Human polynucleoti
22	94.4	34.0	1237	22	AAC77309	Human ORFX ORF2864
23	94.4	34.0	1388	22	AAC91326	Human polynucleoti
24	70.2	25.3	1657	22	ABL02803	Human polynucleoti
25	44.6	16.0	2509	23	ABL02803	Drosophila melanog
26	44.6	16.0	2552	23	ABL20031	Drosophila melanog
27	39.4	14.2	52298	16	AAT51411	Mycobacteriophage
28	39.4	14.2	514	21	AAQ47357	L5 mycobacteriopha
29	39.2	14.1	514	21	AAC50296	Arabidopsis thali
30	37.8	13.6	1446	22	AAH52003	Mycobacterium tube
31	37.6	13.5	9750	23	ABL02802	Drosophila melanog
32	37.6	13.5	17741	22	ABL20030	Drosophila melanog
33	37.4	13.4	51	22	AAH89246	Drosophila melanog
34	37.4	13.4	2041	22	AAH18056	Human cDNA sequenc
35	37.2	13.4	3768	21	AAC81955	Human cDNA sequenc
36	37.2	13.4	3995	22	AAI68201	Human osteoclast e
37	37	13.3	783	22	AAH07715	Human cDNA clone (
38	35.8	12.9	51	22	AAH89247	Human structural p
39	35.4	12.7	1675	23	ABL14443	Drosophila melanog
40	35.4	12.7	1792	23	ABL14347	Drosophila melanog
41	35.4	12.7	6720	23	ABL14342	Drosophila melanog
42	35.4	12.7	6720	23	ABL14346	Drosophila melanog
43	35.2	12.7	538	20	AAZ58276	Zea mays SCLBr pro
44	35.2	12.7	845	21	AAZ49560	Maize MLO1 protein
45	35	12.6	596	21	AAC40660	Arabidopsis thalia

ALIGNMENTS

```
RESULT 1
AAC76864

ID   AAC76864 standard; cDNA; 906 BP.
XX
AC   AAC76864;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2419 polynucleotide sequence SEQ ID NO:4837.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
OS   Homo sapiens.
XX
CC   WO200058473-A2.
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
```

XX
PR   31-MAR-1999; 99US-0127607.
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,   Leach M;
XX
PI
DR   WPI; 2000-603362/57.
DR   P-PSDB; AAB42655.
XX
PS   Claim 5; Page 4020-4021; 5507pp; English.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 906 BP; 302 A; 208 C; 235 G; 160 T; 1 other;

Query Match   95.5%; Score 265.6; DB 21; Length 906;
Best Local Similarity   96.4%; Pred. No. 1e-66;
Matches 268; Conservative   0; Mismatches 10; Indels   0; Gaps   0;

Qy   1 gccaacacgcantccgacgacagtgcagccatggtcattgcagagatgcncaaagtcaat 60
             |||||||||| ||||||| ||||||||||||||||||||||||||||| | ||||||||||
Db   415 gccaacacgcatgccgacgacagtgcagccatggtcattgcagagatgctcaaagtcaat 474

Qy   61 gagcaacatcacaaacgtaaaacgtaaaacgtcgagtccaactcataacggaaaagggcc 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   475 gagcaacatcaccaaacgtaaaacgtcgagtccaactcataacggaaaagggatcctggc 534

Qy   121 atcatgagagctccagcacacaacgtgtcaccggagctgcgcgtttcataacacagagg 180
             ||||||||||||||||||||||||||||||||||||||||||||| || |||||||||||
Db   535 atcatgagagctccagcacacaacgtgtctcaccggagctcgtttccataacacagagg 594

Qy   181 cacatcatggccaggccaggtggaaattggaaattgtcaagctcnctgaagtgaagcacacgacg 240
             |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   595 cacatcatgggccaggccaggtggaaatggaaatgtcaagctgctgaaggagagaacacgacg 654

Qy   241 ctnctgaggctgggntaccattttnaactcccaggacc 278
             || ||||||||||| |||||| | |||||||||||||||
Db   655 ctgctgaggctgggataccattttgaactcccaggacc 692

RESULT 2
AAS31144
ID   AAS31144 standard; cDNA; 2281 BP.
XX

AC   AAS31144;
XX
DT   04-DEC-2001 (first entry)
XX
DE   Human diagnostic and therapeutic polynucleotide (DITHP) #159.
XX
KW   Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW   cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW   acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW   respiratory disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200162927-A2.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-US06059.
XX
PR   24-FEB-2000; 2000US-0184693.
PR   24-FEB-2000; 2000US-0184697.
PR   24-FEB-2000; 2000US-0184698.
PR   24-FEB-2000; 2000US-0184768.
PR   24-FEB-2000; 2000US-0184769.
PR   24-FEB-2000; 2000US-0184770.
PR   24-FEB-2000; 2000US-0184771.
PR   24-FEB-2000; 2000US-0184772.
PR   24-FEB-2000; 2000US-0184773.
PR   24-FEB-2000; 2000US-0184774.
PR   24-FEB-2000; 2000US-0184776.
PR   24-FEB-2000; 2000US-0184777.
PR   24-FEB-2000; 2000US-0184797.
PR   24-FEB-2000; 2000US-0184813.
PR   24-FEB-2000; 2000US-0184837.
PR   24-FEB-2000; 2000US-0184841.
PR   24-FEB-2000; 2000US-0184844.
PR   24-FEB-2000; 2000US-0185213.
PR   12-MAY-2000; 2000US-0185216.
PR   24-FEB-2000; 2000US-0203785.
PR   15-MAY-2000; 2000US-0204226.
PR   16-MAY-2000; 2000US-0204525.
PR   16-MAY-2000; 2000US-0204821.
PR   16-MAY-2000; 2000US-0204908.
PR   16-MAY-2000; 2000US-0205232.
PR   17-MAY-2000; 2000US-0204815.
PR   17-MAY-2000; 2000US-0204863.
PR   17-MAY-2000; 2000US-0205221.
PR   17-MAY-2000; 2000US-0205285.
PR   17-MAY-2000; 2000US-0205286.
PR   17-MAY-2000; 2000US-0205287.
PR   17-MAY-2000; 2000US-0205323.
PR   17-MAY-2000; 2000US-0205324.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI   Chen A, D'Sa SA, Anshey S, Dahl CR, Dam TC, Daniels SE;
PI   Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
PI   Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI   Wright RJ, Yap PE, Yu JTY, Bradley DL, Bratcher SR, Chen W;
PI   Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR   WPI; 2001-502867/55.
DR   P-PSDB; AAU19573.
XX
PS   Claim 1; Page 372-373; 522pp; English.
XX
PT   The invention relates to polynucleotides (I) encoding diagnostic and
PT   therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
PT   enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT   -
XX
CC   Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
CC   enzymes, hormones and receptors, useful in growth and development and
CC   and proteins involved in growth and development and receptors. (I) and

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2002, 09:52:18 ; Search time 69.62 Seconds
                                (without alignments)
                                980.840 Million cell updates/sec

Title:
Perfect score: 278
Sequence: 1 gccaacacgcantccgacga.........catttnaactcccaggacc 278
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

Database :   Issued_Patents_NA:*
               1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
               2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
               3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
               4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
               5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39.4	14.2	52297	4	US-09-426-436-1	Sequence 1, Appli
2	39.4	14.2	52297	4	US-08-705-557-1	Sequence 1, Appli
3	35.2	12.7	538	4	US-09-183-959-13	Sequence 13, Appli
4	34.8	12.5	687	5	PCT-US94-04361-25	Sequence 25, Appl
5	34.6	12.4	681	5	PCT-US94-04361-27	Sequence 27, Appl
6	33.8	12.2	2370	2	US-08-838-219B-19	Sequence 19, Appl
7	33.8	12.2	2370	3	US-09-233-336A-19	Sequence 19, Appl
8	33.8	12.2	2370	4	US-09-233-752A-19	Sequence 19, Appl
9	33.8	12.2	2370	4	US-09-402-036-19	Sequence 19, Appl
10	33.8	12.2	2403	1	US-08-471-033-30	Sequence 30, Appl
11	33.8	12.2	2403	2	US-08-471-044-30	Sequence 30, Appl
12	33.8	12.2	2403	2	US-08-463-483A-30	Sequence 30, Appl
13	33.8	12.2	2403	2	US-08-471-046A-30	Sequence 30, Appl
14	33.8	12.2	2403	2	US-08-470-566B-30	Sequence 30, Appl
15	33.8	12.2	2403	2	US-08-838-219B-7	Sequence 7, Appl
16	33.8	12.2	2403	2	US-08-469-334-30	Sequence 30, Appl
17	33.8	12.2	2403	3	US-09-300-529-30	Sequence 30, Appl
18	33.8	12.2	2403	3	US-09-233-336A-7	Sequence 7, Appl
19	33.8	12.2	2403	4	US-09-233-752A-7	Sequence 7, Appl
20	33.8	12.2	2403	4	US-09-402-036-7	Sequence 7, Appl
21	33.5	12.1	688	5	PCT-US94-04361-26	Sequence 26, App
22	33.4	12.0	777	3	US-09-233-336A-20	Sequence 20, Appl
23	33.4	12.0	2241	2	US-08-838-219B-20	Sequence 20, Appl
c 23	33.4	12.0	2241	2	US-08-838-219B-20	Sequence 20, App
24	33.4	12.0	2241	3	US-09-233-336A-20	Sequence 20, Appl
25	33.4	12.0	2241	4	US-09-233-752A-20	Sequence 20, Appl
26	33.4	12.0	2241	4	US-09-402-036-20	Sequence 20, Appl
27	33.4	12.0	2846	4	US-09-613-182-5	Sequence 5, Appli
c 28	33.4	12.0	50341	1	US-08-247-901C-1	Sequence 1, Appli
c 29	33.4	12.0	50341	2	US-09-075-904-1	Sequence 1, Appli
30	33	11.9	349	1	US-08-275-225-13	Sequence 13, Appli
31	33	11.9	2030	1	US-08-530-950-1	Sequence 1, Appli
32	33	11.9	2030	3	US-08-888-429A-1	Sequence 1, Appli
33	33	11.9	2030	4	US-08-149-879-1	Sequence 1, Appli
34	31.8	11.4	2576	1	US-08-471-033-35	Sequence 35, Appl
35	31.8	11.4	2576	2	US-08-471-044-35	Sequence 35, Appl
36	31.8	11.4	2576	2	US-08-463-483A-35	Sequence 35, Appl
37	31.8	11.4	2576	2	US-08-471-046A-35	Sequence 35, Appl
38	31.8	11.4	2576	3	US-09-300-529-35	Sequence 35, Appl
39	31.8	11.4	2576	3	US-09-470-566B-35	Sequence 35, Appl
40	31.8	11.4	2576	1	US-08-469-334-35	Sequence 35, Appl
41	31.8	11.4	2655	1	US-09-300-529-35	Sequence 35, Appl
42	31.8	11.4	2655	1	US-08-471-033-26	Sequence 26, Appl
43	31.8	11.4	2655	2	US-08-471-044-17	Sequence 17, Appl
44	31.8	11.4	2655	2	US-08-471-044-26	Sequence 26, Appl
45	31.8	11.4	2655	2	US-08-463-483A-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-426-436-1/c
Sequence 1, Application US/09426436
Patent No. 6225066
GENERAL INFORMATION:
  APPLICANT: William R. Jacobs, Jr.
  APPLICANT: Barry R. Bloom
  APPLICANT: Graham F. Hatfull
  TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC
  TITLE OF INVENTION: REPORTER MYCOBACTERIOPHAGES
  NUMBER OF SEQUENCES: 1
  CORRESPONDENCE ADDRESS:
   ADDRESSEE: Amster, Rothstein & Ebenstein
   STREET: 90 Park Avenue
   CITY: New York
   STATE: New York
   COUNTRY: U.S.A.
   ZIP: 10016
  COMPUTER READABLE FORM:
   MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
   COMPUTER: IBM PC Compatible
   OPERATING SYSTEM: MS-DOS
   SOFTWARE: Word Processor (ASCII)
  CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/426,436
   FILING DATE:
   CLASSIFICATION:
  PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US/08/705,557
   FILING DATE:
   APPLICATION NUMBER: US/08/057,531
   FILING DATE:
   APPLICATION NUMBER: 07/833,431
   FILING DATE: February 7, 1992
  ATTORNEY/AGENT INFORMATION:
   NAME: Pasqualini, Patricia A.
   REGISTRATION NUMBER: 34,894
   REFERENCE/DOCKET NUMBER: 96700/238
  TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 697-5995
   TELEFAX: (212) 286-0854 or 286-0082
   TELEX: TWX 710-581-4766
  INFORMATION FOR SEQ ID NO: 1:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 52297
   TYPE: nucleotide
   STRANDEDNESS: single
   TOPOLOGY: linear
   MOLECULE TYPE:
   DESCRIPTION: phage genome sequence

```
;    HYPOTHETICAL:  no
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:  not applicable.
;    ORIGINAL SOURCE:
;      ORGANISM:  mycobacteriophage L5
;      STRAIN:  not applicable
;      INDIVIDUAL ISOLATE:  L5
;      DEVELOPMENTAL STAGE:  not applicable
;      HAPLOTYPE:  not applicable
;      TISSUE TYPE:  not applicable
;      CELL TYPE:  not applicable
;      CELL LINE:  not applicable
;      ORGANELLE:  not applicable
;    IMMEDIATE SOURCE:  mycobacteriophage L5 particles
;    POSITION IN GENOME:  entire genome
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;    OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Hatfull and Sarkis
;      TITLE:  DNA Sequence, Structure and Gene
;      TITLE:  Expression of Mycobacteriophage L5:
;      TITLE:  A Phage System for Mycobacterial
;      JOURNAL:  Genetics
;      JOURNAL:  Molecular Microbiology
;      VOLUME:  7
;      PAGES:  395-405
;      DATE:  1993
;    US-09-426-436-1

Query Match.
Best Local Similarity   14.2%;
Matches  109;  Conservative   47.8%;  Pred. No. 0.033;  Mismatches  119;  Indels   0;  Gaps   0;

                                                           Score 39.4;  DB 4;  Length 52297;

Qy     27 agccatggtcattgcagagatgcncaaagtcaatgagcacatcaccaacgtaaacgtcga  86
          || ||||| | |||||||| ||  ||| ||| || || ||||| ||||| ||||| |||
Db  46884 AGGGTTGGACACCACCGACCTGCCCGACCGCCATCCAGTGGGTCACCTACGTCACCGGCAA  46825

Qy     87 gtccaacttcataacggaaagggggatcctggccatcatgtgagagcttccagcacaacac  146
          | ||  || || ||||||||| |||||||  |||||| ||||||| | ||||| |||| ||
Db  46824 GTACGTTGCCGCACAAGCGTTGTTGCTGCACGGGCCCTTGCACTGCAAGAC  46765

Qy    147 ggtgcttcacggagctgcggttcataaccagaggcacatcatggcagccaggtggaat  206
          ||||  | ||  |  | | ||| |   |||||| | ||| ||||| ||||  | |||
Db  46764 GGTGTACGCCAGCCTCACGGGCGGGTCAGGGGGATCTACGTCTTACGGCGGTGCCCGCCACAT  46705

Qy    207 ggagattgtcaagctnctgaaggagaacacgacgacgctnctgaggctggg  254
          ||| | ||  || | |||  |||  |  | | | | | | | | |||||
Db  46704 CGAGAGGCTCCAATTCCTCTCGGTCGATCATGCACCGCGAGATGATTGCG  46657


RESULT   2
US-08-705-557-1/c
Sequence 1, Application US/08705557
Patent No. 6300061
;    GENERAL INFORMATION:
;      APPLICANT:  William R. Jacobs, Jr.
;      APPLICANT:  Barry R. Bloom
;      APPLICANT:  Graham F. Hatfull
;      TITLE OF INVENTION:  MYCOBACTERIAL SPECIES-SPECIFIC
;      TITLE OF INVENTION:  REPORTER MYCOBACTERIOPHAGES
;      NUMBER OF SEQUENCES:  1
;      CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Amster, Rothstein & Ebenstein
;        STREET:  90 Park Avenue
;        CITY:  New York
;        STATE:  New York
;        COUNTRY:  U.S.A.
;        ZIP:  10016
;      COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/705,557
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/057,531
;      FILING DATE:
;      APPLICATION NUMBER:  07/833,431
;      FILING DATE:  February 7, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pasqualini, Patricia A.
;      REGISTRATION NUMBER:  34,894
;      REFERENCE/DOCKET NUMBER:  96700/238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;    INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  52297
;      TYPE:  nucleotide
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;    HYPOTHETICAL:  no
;    DESCRIPTION:  phage genome sequence
;    ANTI-SENSE:  no
;    FRAGMENT TYPE:  not applicable.
;    ORIGINAL SOURCE:
;      ORGANISM:  mycobacteriophage L5
;      STRAIN:  not applicable
;      INDIVIDUAL ISOLATE:  L5
;      DEVELOPMENTAL STAGE:  not applicable
;      HAPLOTYPE:  not applicable
;      TISSUE TYPE:  not applicable
;      CELL TYPE:  not applicable
;      CELL LINE:  not applicable
;      ORGANELLE:  not applicable
;    IMMEDIATE SOURCE:  mycobacteriophage L5 particles
;    POSITION IN GENOME:  entire genome
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;    OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:  Hatfull and Sarkis
;      TITLE:  DNA Sequence, Structure and Gene
;      TITLE:  Expression of Mycobacteriophage L5:
;      TITLE:  A Phage System for Mycobacterial
;      JOURNAL:  Genetics
;      JOURNAL:  Molecular Microbiology
;      VOLUME:  7
;      PAGES:  395-405
;      DATE:  1993
;    US-08-705-557-1

Query Match
Best Local Similarity   14.2%;
Matches  109;  Conservative   47.8%;  Pred. No. 0.033;  Mismatches  119;  Indels   0;  Gaps   0;

                                                           Score 39.4;  DB 4;  Length 52297;

Qy     27 agccatggtcattgcagagatgcncaaagtcaatgagcacatcaccaacgtaaacgtcga  86
          || ||||| | |||||||| ||  ||| ||| || || ||||| ||||| ||||| |||
Db  46884 AGGGTTGGACACCACCGACCTGCCCGACCGCCATCCAGTGGGTCACCTACGTCACCGGCAA  46825

Qy     87 gtccaacttcataacggaaagggggatcctggccatcatgtgagagcttccagcacaacac  146
          | ||  || || ||||||||| |||||||  |||||| ||||||| | ||||| |||| ||
Db  46824 GTACGTTGCCGCACAAGCGTTGTTGCTGCACGGGCCCTTGCACTGCAAGAC  46765
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: June 15, 2002, 09:47:03 ; Search time 2032.22 Seconds
(without alignments)
2862.671 Million cell updates/sec

Title: US-09-445-362B-6
Perfect score: 278
Sequence: 1 gccaacacgcantccgacga.........cattttnaactcccaggacc 278

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326829 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
   10: gb_ro:*
   11: gb_sts:*
   12: gb_sy:*
   13: gb_un:*
   14: gb_vi:*
   15: gb_ba:*
   16: em_ba:*
   17: em_fun:*
   18: em_hum:*
   19: em_in:*
   20: em_mu:*
   21: em_om:*
   22: em_or:*
   23: em_ov:*
   24: em_pat:*
   25: em_ph:*
   26: em_pl:*
   27: em_ro:*
   28: em_sts:*
   29: em_un:*
   30: em_vi:*
   31: em_htg_hum:*
   32: em_htg_inv:*
   33: em_htg_other:*
       em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match %	Length	DB	ID	Description
1	265.6	95.5	187532	9	AC006333	AC006333 Homo sapi
2	264	95.0	1936	6	AX002051	AX002051 Sequence
3	264	95.0	2080	6	AX002052	AX002052 Sequence
4	264	95.0	2268	6	AX002053	AX002053 Sequence
5	261	93.9	279	6	AX002056	AX002056 Sequence
6	216	77.7	168526	2	AC103173	AC103173 Rattus no
7	214.4	77.1	1458	10	AF237628	AF237628 Mus muscu
8	152	54.7	21542	10	AL513217	AL513217 Human DNA
9	152	54.7	171423	2	AC025664	AC025664 Homo sapi
10	150.4	54.1	3881	6	AX329789	AX329789 Sequence
11	150.4	54.1	3881	6	AX331533	AX331533 Sequence
12	150.4	54.1	3881	6	AX331932	AX331932 Sequence
13	150.4	54.1	3881	9	AX335915	AX335915 Sequence
14	150.4	54.1	3881	9	HSAUTAN6A	X54162 Human mRNA
15	150.4	49.5	1760	9	AC016814	AC016814 Mus muscu
16	137.6	49.5	182151	2	AK057852	AK057852 Homo sapi
17	137.6	47.2	518	11	AC092060	AC092060 Homo sapi
18	131.2	47.2	518	11	G56351	G56351 SHGC-101709
19	120.8	43.5	93276	10	AC096239	AC096239 Rattus no
20	118.8	42.7	1572	10	S76831	S76831 Tmod=tropo
21	115.6	41.6	1837	10	AF237629	AF237629 Mus muscu
22	114	41.0	1550	10	AF177170	AF177170 Rattus norv
23	110.4	39.7	1590	10	RNU59240	U59240 Rattus norv
24	109.2	39.3	1353	10	RNU59241	U59241 Rattus norv
25	106	38.1	1067	5	AB052132	AB052132 Gallus ga
26	104.4	37.6	2493	5	L36678	L36678 Gallus dome
27	104	37.4	2491	9	CHKTMOD1A	AF177169 Homo sapi
28	101.8	36.6	1780	9	BC002660	BC002660 Homo sapi
29	101.8	36.6	2665	9	HUMTROMOD	M77016 Human tropo
30	94.4	34.0	1584	9	AF177171	AF177171 Homo sapi
31	94.4	34.0	2072	9	AF237631	AF237631 Homo sapi
32	94.4	34.0	2142	9	BC020542	BC020542 Homo sapi
33	89.6	32.2	1806	4	AF237632	AF237632 Sus scrof
34	88.4	31.8	1231	5	AF165215	AF165215 Gallus ga
35	88	31.7	1442	5	AF165216	AF165216 Danio rer
36	86.8	31.2	1001	5	AB052717	AB052717 Gallus ga
37	84.8	30.5	3540	10	AF237630	AF237630 Mus muscu
38	83.2	29.9	1764	10	AF177172	AF177172 Mus muscu
39	73.4	26.4	1161	10	AF177174	AF177174 Mus muscu
40	70.2	25.3	1260	9	BC017810	BC017810 Homo sapi
41	70.2	25.3	1267	9	AF177173	AF177173 Homo sapi
42	68.6	24.7	872	10	AF287744	AF287738S7 Mus muscu
43	68.6	24.7	1235	9	AF165217	AF165217 Homo sapi
44	68	24.5	3551	9	AL137543	AL137543 Homo sapi
45	61.2	22.0	162863	2	AC096430	AC096430 Rattus no

ALIGNMENTS

RESULT 1
AC006333
LOCUS   AC006333   187532 bp   DNA   linear   PRI 30-SEP-2000
DEFINITION Homo sapiens BAC clone RP11-390E23 from 7, complete sequence.
ACCESSION   AC006333
VERSION   AC006333.3 GI:5523811
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 187532)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE   Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
REFERENCE   2 (bases 1 to 187532)
AUTHORS   Rohlfing,T., Bauer,C. and Ketterman,M.
TITLE   The sequence of Homo sapiens BAC clone RP11-390E23
JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 187532)
AUTHORS   Waterston,R.H.
TITLE   Direct Submission

```
JOURNAL   Submitted (09-JAN-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 4 (bases 1 to 187532)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 5 (bases 1 to 187532)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Jul 17, 1999 this sequence version replaced gi:4454617.
          --------------  Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
          --------------  Summary Statistics
            Center project name: H_NH0390E23
          --------------

          NOTICE:  This sequence may not represent the entire insert of this
          clone.  It may be shorter because we only sequence overlapping
          clone sections once, or longer because we provide a small overlap
          between neighboring data submissions.

          This sequence was finished as follows unless otherwise noted:
          all regions were double stranded, sequenced with an alternate
          chemistry, or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by sequence
          from more than one subclone; and the assembly was confirmed by
          restriction digest.

          MAPPING INFORMATION:
          The sequence of this clone was established as part of a mapping and
          sequencing collaboration between the NHGRI Chromosome 7 Mapping
          Project (Eric D. Green, Director), John D. McPherson in the
          Department of Genetics (Washington University), and the Washington
          University Genome Sequencing Center. For additional information
          about the map position of this sequence, see
          http://www.nhgri.nih.gov/DIR/GTB/CHR7/, send
          mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

          SOURCE INFORMATION:
          The RPCI-11 human BAC library was made from the blood of one male
          donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
          Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
          approach for construction of bacterial artificial chromosome
          libraries.  Genomics 51:1-8.  The clone may be obtained either from
          Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
          and coworkers at the Roswell Park Cancer Institute
          (http://bacpac.med.buffalo.edu)
          VECTOR:  pBACe3.6
          NEIGHBORING SEQUENCE INFORMATION:
          The clone sequenced to the right is a GS1-195F7, 200 bp overlap.
          Actual start is at base position 1 of RP11-390E23; actual end is at
          1716 of GS1-195F7.

          The sequence fidelity is in question at base positions 75394-75555
          due to a non-perfect GA run. The insilico digest fragments are 3570
          for ecorv and 8922 for hindiii compared to 3573 for ecorv and 8997
          hindiii in the real fragment digest. Assembly appears to be correct
          according to the restriction digests.
FEATURES             Location/Qualifiers
     source          1..187532
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-390E23"
                     /clone_lib="RPCI-11"
     repeat_region   2..182
                     /rpt_family="Retroviral"
     repeat_region   190..231
                     /rpt_family="L1"
     repeat_region   327..386
                     /rpt_family="(CA)n"
     repeat_region   392..526
                     /rpt_family="MaLR"
     repeat_region   527..834
                     /rpt_family="Alu"
     repeat_region   836..1227
                     /rpt_family="Alu"
     repeat_region   3024..3337
                     /rpt_family="MaLR?"
     repeat_region   3856..4008
                     /rpt_family="Alu"
     repeat_region   4823..5128
                     /rpt_family="L2"
     repeat_region   5696..5775
                     /rpt_family="Alu"
     repeat_region   6130..6323
                     /rpt_family="MIR"
     misc_feature    6433..6524
                     /note="match to EST AA984773 (NID:g3163298) am90h03.s1"
     repeat_region   6591..6880
                     /rpt_family="L1"
     repeat_region   6925..7231
                     /rpt_family="Alu"
     repeat_region   7258..7561
                     /rpt_family="Alu"
     repeat_region   8101..8622
                     /rpt_family="L1"
     misc_feature    9122..9332
                     /note="match to EST AA716350 (NID:g2728624) zh29d08.s1"
     repeat_region   10506..10842
                     /rpt_family="MER1_type"
     repeat_region   10846..10881
                     /rpt_family="Other"
     repeat_region   10912..10940
                     /rpt_family="(TA)n"
     repeat_region   12699..12908
                     /rpt_family="(TTTTA)n"
     repeat_region   13697..13718
                     /rpt_family="L1"
     repeat_region   13719..14081
                     /rpt_family="(TTTTA)n"
     repeat_region   15249..15401
                     /rpt_family="L1"
     repeat_region   15380..15568
                     /rpt_family="MIR"
     repeat_region   17375..17504
                     /rpt_family="MER1_type?"
     repeat_region   17612..17661
                     /rpt_family="L2"
     repeat_region   17774..17796
                     /rpt_family="L2"
     repeat_region   17905..18016
                     /rpt_family="AT_rich"
     repeat_region   18017..18048
                     /rpt_family="Alu"
     repeat_region   18650..19009
                     /rpt_family="(TAAA)n"
     repeat_region   19010..19133
                     /rpt_family="L2"
     repeat_region   19164..19254
                     /rpt_family="Alu"
     repeat_region   19383..19687
                     /rpt_family="MER1_type"
     repeat_region   20281..20428
                     /rpt_family="L1"
     repeat_region   20429..20456
                     /rpt_family="L1"
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: June 13, 2002, 09:45:35 ; Search time 30.86 seconds
        (without alignments)
        3094.401 Million cell updates/sec

Title: US-09-445-362B-4
Perfect score: 2839
Sequence: 1 MSTFGYRRGLSKYESIDEDE.........SIKQLKRVEVPEALRWEHDL 552

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Total number of hits satisfying chosen parameters: 562222

Database : SPTREMBL.19:*
   1: sp_archea:*
   2: sp_bacteria:*
   3: sp_fungi:*
   4: sp_human:*
   5: sp_invertebrate:*
   6: sp_mammal:*
   7: sp_mhc:*
   8: sp_organelle:*
   9: sp_phage:*
  10: sp_plant:*
  11: sp_rodent:*
  12: sp_virus:*
  13: sp_vertebrate:*
  14: sp_unclassified:*
  15: sp_rvirus:*
  16: sp_bacterial:*
  17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1536	54.1	333	11	Q99PM7	q99pm7 mus musculu
2	679	23.9	348	4	Q96LS4	q96ls4 homo sapien
3	656.5	23.1	351	11	Q9JKK7	q9jkk7 mus musculu
4	656.5	23.1	351	11	P70566	p70566 rattus norv
5	656.5	23.1	359	11	P70567	p70567 rattus norv
6	650.5	22.9	351	4	Q9NZR1	q9nzr1 homo sapien
7	650.5	22.9	351	11	Q9JLH9	q9jlh9 mus musculu
8	650.5	22.9	359	11	Q9ERR9	q9err9 mus musculu
9	644.5	22.7	359	13	Q91006	q91006 gallus gall
10	627.5	22.1	359	4	Q9BUF1	q9buf1 homo sapien
11	614	21.6	352	4	Q9NZR0	q9nzr0 homo sapien
12	613	21.5	352	6	Q9NYL9	q9nyl9 homo sapien
13	609	21.5	352	4	Q9N0Y9	q9n0y9 homo sapien
14	606.5	21.4	345	13	Q9DEA6	q9dea6 gallus gall
15	590	20.8	352	11	Q9JHJ0	q9jhj0 mus musculu
16	561	19.8	345	4	Q9NZQ9	q9nzq9 homo sapien
17	560	19.7	345	11	Q9JLH8	q9jlh8 mus musculu
18	555	19.5	345	4	Q9UKH2	q9ukh2 homo sapien
19	546.5	19.2	343	13	Q9PUQ6	q9puq6 brachydanio
20	541	19.1	343	13	Q9PUQ7	q9puq7 gallus gall
21	517	18.2	333	13	Q9DEA4	q9dea4 gallus gall
22	382.5	13.5	392	5	Q01479	001479 caenorhabdi
23	374	13.2	367	5	046231	046231 drosophila
24	369	13.0	403	5	Q9VA56	q9va56 drosophila
25	369	13.0	401	5	Q95Q88	q95q88 caenorhabdi
26	248.5	8.8	269	4	Q9BV00	q9bv00 homo sapien
27	242.5	8.5	1077	11	Q9JHC1	q9jhc1 mus musculu
28	240.5	8.5	737	4	Q9Y2L7	q9y2l7 homo sapien
29	240.5	8.5	1013	4	Q9NT81	q9nt81 homo sapien
30	228	8.0	1112	4	Q96PY5	q96py5 homo sapien
31	228	8.0	168	11	Q9D386	q9d386 mus musculu
32	226.5	8.0	602	2	Q9AKP3	q9akp3 rickettsia
33	226.5	8.0	1179	12	Q9IL98	q9il98 white spot
34	225.5	7.9	4833	11	Q9QYX6	q9qyx6 mus musculu
35	225.5	7.9	5038	11	Q9QYX7	q9qyx7 mus musculu
36	224.5	7.9	240	11	Q9CUK4	q9cuk4 mus musculu
37	212	7.5	89	4	Q9NT43	q9nt43 homo sapien
38	212	7.5	3394	4	Q9Y6V0	q9y6v0 homo sapien
39	211.5	7.4	560	4	Q9H985	q9h985 homo sapien
40	211.5	7.4	1520	4	O15087	o15087 homo sapien
41	211.5	7.4	1781	4	O15087	o15087 homo sapien
42	211.5	7.4	1890	4	Q9UKX0	q9ukx0 homo sapien
43	211	7.4	2073	4	Q9UKW3	q9ukw3 homo sapien
44	211	7.4	542	4	Q9UKW2	q9ukw2 homo sapien
45	211	7.4	1110	5	O11448	o11448 anagrapha f
46	211	7.4	1110	13	Q91255	q91255 petromyzon

ALIGNMENTS

RESULT 1
Q99PM7

ID   Q99PM7   PRELIMINARY;   PRT;   333 AA.
AC   Q99PM7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CARDIAC LEIOMODIN (FRAGMENT).
GN   LMOD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21218919; PubMed=11316603;
RA   Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;
RT   "Leiomodins: larger members of the tropomodulin (tmod) gene family.";
RL   Genomics 73:127-139(2001).
DR   EMBL; AF237628; AAK00789.1; -.
DR   MGD; MGI:2135672; Lmod2.
FT   NON_TER   1   1
SQ   SEQUENCE   333 AA;   37463 MW; F082494410157A31 CRC64;

Query Match   54.1%; Score 1536; DB 11; Length 333;
Best Local Similarity   90.1%; Pred. No. 1.4e-75;
Matches   301; Conservative   13; Mismatches   16; Indels   4; Gaps   2;

```
Qy    214 TEVNLNNIENITTQTLRFAEALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNV 273
          || |||||||||||||||||| ::|:|||:||||||||||||::||:|||||||||| :|
Db      4 TEFNLNNIENITTQTLSRFEALKENTVVKTFSLANTHADDAAAIAIADMLKVNEHITSV 63

Qy    274 NVESNFITGKGILAIMRALQHNTVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGY 333
          |||||||||||||||||||||:|||||||||:|||||||||||||||||||||||||||| 
Db     64 NVESNFITGKGILAIMRALQHNTVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGY 123

Qy    334 HFELPGPRMSMTSILTRNMDKQRQKRLQEQKQQEGYDGGPNLRTKVWQRGTPSSSPYVSP 393
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  124 HFELPGPRMSMTSILTIRNMDKQRQKRMQEQKQQEGHDGGAALRTKVWQRGTPGSSPYASP 183

Qy  394 RHSPWSSSPKLPKVQTVRSRPLSPVATLPPPPPPPSSORLPPPPPPPPPLPEKK 453
Db  184 RQSPWSSSPKVSKKVHTGRSRPPSPVA--PPPPPPPLPP--HMLPPPPPPAPPLPEKK 239

Qy  454 LITRNIAEVIKQQESAQRALQNGQKKKGKKVKKQPNSILKEIKNSLRSVQEKKMEDSSR 513
Db  240 LITRNIAEVIKQQESAQRALQNGQRKKKGKKVKKQPNNILKEIKNSLRSVQEKKMEDSSR 299

Qy  514 PSTPQRSAHENLMEAIRGSSIKQLKRVEVPEALR 547
Db  300 PSTPQRSVHENLMEAIRGSSIRQLRRVEVPEALR 333
```

```
RESULT 2
ID   Q96LS4                PRELIMINARY;   PRT;   348 AA.
AC   Q96LS4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ25123 FIS, CLONE CBR06154.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Tashiro H., Yamazaki M., Watanabe K., Kunagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Suzuki Y., Hata H.,
RA   Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA   Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA   Nagai K., Isogai T., Sugano S.;
RL   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK057852; BAB71596.1; -.
SQ   SEQUENCE 348 AA; 40194 MW; 6D05E271D460B087 CRC64;

Query Match             23.98;  Score 679;  DB 4;   Length 348;
Best Local Similarity   45.2%;  Pred. No. 2.3e-29;
Matches    141;  Conservative  53;  Mismatches 102;  Indels  16;  Gaps  2;

Qy   72 YWEKESOQKLLEKERLGECGKVAEDKEESEEELIFTESNSEVSEEVVTEEEEEESOEEEE 131
Db    2 YWEKASRRMLEEERVPVTFVKSEEKLNNEIVANKRESKGSSNIQETDEDEEEEEDDDDDD 61

Qy  132 EDSDEEERTIETAKGINGTVNYDSVNSDNS----RPKIFKSQIENILNLTNGSNGRNTE-- 185
Db   62 EGEDDGESEEETNREEEGKAKKEQIRNCENNCCQQVTDKAFKEQRDRPEAOEQSEKKISKLD 121

Qy  186 -----------SPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTRFAEA 235
Db  122 PKKLALDTSFLKVSTRPSGNQTDLDGSLRRVRKNDPDMKELNLNLNIENIPKEMLLDFVNA 181

Qy  236 LKDNTVVKTFSLANTHADDSRAMAIAEMLKANEHITNVNVESNFITGKGILAIMRALQHN 295
Db  182 MKKNKHIKTFSLANVGADENVAFALANMLRENRSITTLNIESNFITGGIVAIMRCLQFN 241

Qy  296 TVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSILTRNMDKQ 355
Db  242 ETLTELRFHNQRHMLGHHAEMEIARLLKANNTLLKMGYHFELPGPRMVVVNLLTRNQDKQ 301

Qy  356 ROKRLQEEKQQE 367
Db  302 ROKRQEEOKQQO 313
```

```
RESULT 3
ID   Q9JKK7                PRELIMINARY;   PRT;   351 AA.
AC   Q9JKK7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NEURAL TROPOMODULIN N-TMOD.
DE   TMOD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Conley C.A., Fowler V.M.;
RT   "Identifying novel tropomodulin isoforms.";
RL   Mol. Biol. Cell 9:18A-18A(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Conley C.A., Almenar-Queralt A., Fowler V.M.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF237629; AAF45297.1; -.
MGD; MGI:1355335; Tmod2.
SQ   SEQUENCE 351 AA; 39510 MW; D01FDDEE185C828F CRC64;

Query Match             23.18;  Score 656.5;  DB 11;  Length 351;
Best Local Similarity   39.9%;  Pred. No. 3.7e-28;
Matches    148;  Conservative  72;  Mismatches 116;  Indels  35;  Gaps  8;

Qy    6 YRRGLSKYESIDEDELLASLSAEELKELERELEDIEPD-RNLPVGLRQKSLTEKTPTGTF 64
Db    5 FQKGLEKYKNIDEDELIGKLSEEELKQLENVLDDLDPESATLPAGFRQKDQTQKAATGPF 64

Qy   65 SREALMAYWEKESOQKLLEKERLGECGKVAEDKEESEEELIFT-ESNSEV----SEEVYTE 119
Db   65 DREHLLMYLEKE------------------------ALEQKDREDFVPFTGEKKGRVFIPKEKPVETR 108

Qy  120 EEEEESOEEEEEE-----DSDEEERTIETAKGINGTVNYDSVNSDNSKPKIFKSQIENINL 175
Db  109 KEEKVTLDPELEEEALASASDTELYDLAAVLGVHNLLNNPKFPDEFTTNGQGRKGPVRNV-- 166

Qy  176 TNGSNGRNT-ESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTRFAE 234
Db  167 VKGEAKPVFEEPP-----NPTNVEASLQOMKANDPSLQEVNLNNIKNIPIPLKEFAK 220

Qy  235 ALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNVNVESNFITGKGILAIMRALOH 294
Db  221 SLETNTHVKKFSLAATRSNDPVALAFAEMLKVNKILKSLNVESNFITGCTILALVEALRE 280

Qy  295 NTVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSILTRNMDK 354
Db  281 NDTLTEIKIDNQROQLGTAVEMEIAQMLEENSRILKFGYQFTKQGPRTRVAAAITKNNDL 340

Qy  355 QROKRLQEEKQQ 365
Db  341 VRKKRVEGDRR 351
```

```
RESULT 4
ID   P70566                PRELIMINARY;   PRT;   351 AA.
AC   P70566;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   N-TROPOMODULIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE=97041709; PubMed=8886980;
```

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:             June 13, 2002, 09:45:55 ; Search time 13.45 Seconds
                    (without alignments)
                    1589.085 Million cell updates/sec

Title:              US-09-445-362b-4
Perfect score:      2839
Sequence:           1 MSTFGYRRGLSKYESIDEDE...........SIKQLKRVEVEPEALRWEHDL 552

Scoring table:      BLOSUM62
                    Gapop 10.0 , Gapext 0.5

Searched:           105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:   105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	AC	Description
1 *	914	32.2	572	1	LMD1_HUMAN	P29536	homo sapien
2	650.5	22.9	359	1	TMOD_MOUSE	P49813	mus musculu
3	631.5	22.2	359	1	TMOD_HUMAN	P28289	homo sapien
4	209	7.4	2004	1	MOZ_HUMAN	Q92794	homo sapien
5	199	7.0	1953	1	BNI1_YEAST	P41832	saccharomyc
6	193.5	6.8	1375	1	BNR1_YEAST	P40450	saccharomyc
7	189.5	6.7	1790	1	SEPA_EMENI	P78621	emericella
8	185.5	6.5	474	1	VP61_NPVOP	O10270	orgyia pseu
9	182	6.4	485	1	SSGP_VOLCA	P21997	volvox cart
10	182	6.4	802	1	NAB3_YEAST	P38996	saccharomyc
11	182	6.4	1362	1	BRD4_HUMAN	O60885	homo sapien
12	181	6.4	543	1	VP61_NPVAC	O03209	autographa
13	180.5	6.4	1248	1	DIA1_HUMAN	O60610	homo sapien
14	180.5	6.4	2517	1	NCR2_HUMAN	Q9Y618	h nuclear r
15	178.5	6.3	415	1	ACRO_PIG	P08001	sus scrofa
16	178	6.3	384	1	VASP_CANFA	P50551	canis famil
17	177.5	6.2	2472	1	NCR2_MOUSE	Q9WU42	mus musculu
18	176.5	6.2	431	1	ACRO_RABIT	P48038	oryctolagus
19	174.5	6.1	421	1	ACRO_HUMAN	P10323	homo sapien
20	173.5	6.1	1101	1	DIA2_HUMAN	O60879	homo sapien
21	172	6.1	440	1	G3PT_MOUSE	Q64467	mus musculu
22	172	6.1	1206	1	FM14_MOUSE	Q05859	mus musculu
23	171	6.0	380	1	VASP_HUMAN	Q05552	homo sapien
24	171	6.0	771	1	YJC0_YEAST	P47068	saccharomyc
25	171	6.0	1230	1	ST20_CANAL	Q92212	candida alb
26	170.5	6.0	451	1	MYBH_DICDI	P34127	dictyosteli
27	165	5.8	141	1	YPRO_OWFFU	P21260	owenia fusi
28	165	5.8	1091	1	DIA_DROME	P48608	drosophila
29	164	5.8	1468	1	FMN1_MOUSE	Q05860	mus musculu
30	164	5.8	1781	1	AKAC_HUMAN	Q02952	homo sapien
31	163.5	5.8	2774	1	MAPA_RAT	P34926	rattus norv
32	163	5.8	501	1	WASL_RAT	O08816	rattus norv
33	163	5.7	1213	1	FMN_CHICK	O05858	gallus gall
34	163	5.7	1574	1	SYJ1_RAT	Q62910	rattus norv
35	163	5.7	1875	1	MLP1_YEAST	Q02455	saccharomyc
36	162.5	5.7	505	1	WASL_HUMAN	O00401	homo sapien
37	162	5.7	505	1	WASL_BOVIN	Q95107	bos taurus
38	162	5.7	3924	1	ANK2_HUMAN	Q01484	homo sapien
39	161.5	5.7	3726	1	ABF1_MOUSE	Q61329	mus musculu
40	161	5.7	681	1	MP10_HUMAN	Q00566	homo sapien
41	161	5.7	1018	1	FNBA_STAAU	P14738	staphylococ
42	160.5	5.7	487	1	EBN2_EBV	P12978	epstein-bar
43	160.5	5.7	304	1	CEC1_CAEEL	P34618	caenorhabdi
44	159	5.6	2442	1	CBP_HUMAN	Q92793	homo sapien
45	158.5	5.6	379	1	YP85_CAEEL	Q09442	caenorhabdi

ALIGNMENTS

RESULT 1
LMD1_HUMAN

```
ID   LMD1_HUMAN     STANDARD;      PRT;    572 AA.
AC   P29536;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Leiomodin 1 (Leiomodin, muscle form) (64 kDa autoantigen D1) (64 kDa
DE   autoantigen 1D) (64 kDa autoantigen 1D3) (Thyroid-associated
DE   ophthalmopathy autoantigen) (Smooth muscle leiomodin) (SM-Lmod).
GN   LMOD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thyroid;
RX   MEDLINE=91225220; PubMed=2026759;
RA   Dong Q., Ludgate M., Vassart G.;
RT   "Cloning and sequencing of a novel 64-kDa autoantigen recognized by
RT   patients with autoimmune thyroid disease.";
RL   J. Clin. Endocrinol. Metab. 72:1375-1381(1991).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94451105; PubMed=10520227;
RA   Conley C.A., Fowler V.M.;
RT   "Localization of the human 64kD autoantigen D1 to myofibrils in a
RT   subset of extraocular muscle fibers.";
RL   Curr. Eye Res. 19:313-322(1999).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=21248187; PubMed=11350761;
RA   Conley C.A.;
RT   "Leiomodin and tropomodulin in smooth muscle.";
RL   Am. J. Physiol. 280:C1645-C1656(2001).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=21218919; PubMed=11318603;
RA   Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;
RT   "Leiomodins: larger members of the tropomodulin (Tmod) gene family.";
RL   Genomics 73:127-139(2001).
CC   -!- SUBCELLULAR LOCATION: CYTOSKELETON.
CC   -!- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON
CC       AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT
CC       LOW LEVEL IN THYROID.
CC   -!- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.
CC   -!- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.
```

```
CC
DR   EMBL; X54162; CAA38101.1; -.
DR   PIR; S26815; S26815.
DR   MIM; 602715; -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
DR   SMART; SM00246; WH2; 1.
KW   Antigen; Repeat; Cytoskeleton.
FT   DOMAIN       137    265       8 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT       137    152       1.
FT   REPEAT       153    168       2.
FT   REPEAT       169    184       3.
FT   REPEAT       185    200       4.
FT   REPEAT       201    216       5.
FT   REPEAT       217    232       6.
FT   REPEAT       233    248       7.
FT   REPEAT       249    265       8.
FT   DOMAIN       480    499       5 X 4 AA APPROXIMATE TANDEM REPEATS.
FT   DOMAIN       249    265
SQ   SEQUENCE    572 AA;  63737 MW;  D4B42F8052DE94 CRC64;

Query Match         32.2%;   Score 914;   DB 1;   Length 572;
Best Local Similarity   34.2%;   Pred. No. 8.3e-35;
Matches  220;   Conservative  91;   Mismatches 135;   Indels 198;   Gaps 11;

Qy     30 LKELERELEDIEPDRNLPVGLRQKSLTEKTPTGTFSREALMAYWEKESQKLLEKE----- 84
          ::|:::||| ::||::| ||::::|| :::::::||:: :||||::||:::|||:
Db      1 MEELELEKDLVVDPDGSVPVGLRQRNQTEKQSTGVYNREAMLNFCEKETKKLMQREMSMDE 60

Qy     85 ------------------------RLGECGK------------------------ 91
                                  ||:|:||
Db     61 SKQVETKTDAKNGQERGRDASKRALGPRRNSDLGKEPKRGGLKKSFSRDREAGGKSGEK 120

Qy     92 ----------VAEDKEES------------EEELIFTESNSEVS------------- 113
                     |:|:|:|              ||::||::||::
Db    121 PKEEKIIRGIDKGRVRAAVDKKEAGKDGRGEERAVATKKEEEKKGGDRNTGLSRDKDKKR 180

Qy    114 --EEVYTEEEEESQEEEEEEDSDEEERTIETAKG---------INGTVNYDSVN-- 157
            ::| ||||||||:||||||:||||||||::              |::||::|:
Db    181 EEMKEEVAKKEDDEKVKGERRNTDTRKEGEKMKRAGGNTDMKKEDEKVKRGTGNTDTKKDD 240

Qy    158 ----------SDNSKPKIFKSQIENINLNINGNGRNTESPAAIHPCGNPTVIEDAL 203
                    ::|:||| :| ::|::|:: || ::||::| ||| ||||||||||
Db    241 EKVKKKNEPLHEKEAKDDSKTKTPEKQ-----TPSGPTKPSEGPAKVEEEAAPSIFDEPL 294

Qy    204 DKIKSNDPDTTEVNLNNIENITTQTLTRFAEALKDNTVVKTFSLANTHADDSAAMATAEM 263
          |:||:|:||:||::||::| :|:::||:|||||||||||||||:|||:||::||||||||
Db    295 ERVKNNDPEMTEVNVNNNSDCITNEILVRFTEALEFNTVVKLFALANTRADDHVAFAIAIM 354

Qy    264 LKANEHITNVNVESNFITGKGILAIMRALQHNTVLTELRFHNQRHIMGSQVEMEIVKLLK 323
          |||| ::| |||::||:|||| |||||||||||| ||||||||||||||||||||:|||
Db    355 LKANKTITSLNLDSNHITGKGILAIFRALLQNNTLTELRFHNQRHICGGKTEMEIAKLLK 414

Qy    324 ENTTLLRLGYHFELPGPRMSMTSILTTRNMDKQRQKRLQEEKQOEGYDGGPNLRTKVWORG 383
          |||:|:||||||||||||||||:|| |||:||:||:|| ||:||:|:| ||:|: |:|:
Db    415 ENTSLLKLGYHFELAGPRMTVINLLSRNMDKQRQKRLQEORQAOEAKGEKKDLLEVPKAG 474

Qy    384 TPSSSPYVSPRHSPWSSSPKLPKKVQTVRSRPLSPVATLPPPPPPPPPSSQRLPPPPP 443
          :                |  || :::|:| || |::: ||||||||:|||||
Db    475 AVAKG---SPKPSQPSPK-------------------PSPKNSPKKGGAPAAPP 507

Qy    444 PPPPPLPEKKLITRNIAEVIKQQESAQRALQNGQOKKKKGGKVKKQPNSILKEIKNSLRSV 503
          |||||||||| ||::||:|::|:|::||:|::|::|:::||::||::| ::||::||::||
Db    508 PPPPL--------------------------------------APPLIMENLKNSLSPA 529

Qy    504 QEKKMEDSSRPSTPQRSAHENLMEAIRGSSIKQLKRVEVPEALR 547
          :::|                ::|: |:: :::|||||:|::|:|
Db    530 TQRKMGDKVLPA-OEKNSRDQLLAAIRSSNLKQLKKVEVPKLLQ 572
```

```
RESULT 2
TMOD_MOUSE
ID   TMOD_MOUSE              STANDARD;     PRT;    359 AA.
```

```
AC   P49813;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TMOD OR TMOD1.
GN   Tropomodulin.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95154574; PubMed=7851652;
RA   Ito M., Swanson B., Sussman M.A., Kedes L., Lyons G.;
RT   "Cloning of tropomodulin cDNA and localization of gene transcripts
RT   during mouse embryogenesis.";
RL   Dev. Biol. 167:317-328(1995).
CC   -!- FUNCTION: TROPOMYOSIN-REGULATING PROTEIN. IT MAY MODULATE THE
CC       ASSOCIATION OF TROPOMYOSIN WITH THE SPECTRIN-ACTIN COMPLEX IN
CC       THE ERYTHROCYTE MEMBRANE SKELETON, AND THUS THE VISCOELASTIC
CC       PROPERTIES OF ERYTHROCYTES. IT BINDS TO THE END OF ERYTHROCYTE
CC       TROPOMYOSIN AND BLOCKS HEAD-TO-TAIL ASSOCIATION OF TROPOMYOSIN
CC       ALONG ACTIN FILAMENTS (BY SIMILARITY).
DR   EMBL; S76831; AAB33388.1; -.
DR   MGD; MGI:98775; Tmod1.
FT   DOMAIN        39    138       TROPOMYOSIN-BINDING (POTENTIAL).
SQ   SEQUENCE    359 AA;  40484 MW;  5EE340B83048823D CRC64;

Query Match         22.9%;   Score 650.5;   DB 1;   Length 359;
Best Local Similarity   40.1%;   Pred. No. 3.7e-23;
Matches  150;   Conservative  66;   Mismatches 111;   Indels 47;   Gaps 8;

Qy      6 YRGLSKYESIDEDELLASLSAEELKELEDERELELEDIEPDRN-LPVGLRQKSLTEKTPTGTF 64
          ||::| ||::| ||:||::|||| |:||:|:||:|||||||| ||||||||:|||:|||| |
Db      3 YRRELEKYRDLDEDEILGALTEEELRTLENELDELDPDNALLPAGLRQKDQTTKAPTGPF 62

Qy     65 SREALMAYWEKESQKLLEKERL---GE-CGKVAEDKEESSEEELIFTESNSEVSEEVYTE 119
          |||||||||||||||||||:|     |: ||:::||::|||||||::||:||::|| :|
Db     63 KREELLDHLEKQAKEFKDREDLVPYTGEKRGVVWVPKQKPMDPVL-----ESVTLE 113

Qy    120 EEEEESQEEEEEEDSDE------------EERTIETAKGINGTVNYDSVSNDSNSKPKIFKS 168
          ||||:|:|||||:||:               |:||:|:| |::|::|:|:::|||:||:
Db    114 PELEEALANASDAELCDIAAILGMHTLMSNQQYQALGSSSIVNKEGLNS-VIKPTQYKP 172

Qy    169 QIENINLNLTNGSNGRNTESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQT 228
          ::|:|::|::|::|||||||||||||||||||||||::||:|::|:||||:||::|: |
Db    173 VPD-----------EEP-------NSTDVEETLERIKNNDPELEEVNLNNIRNIPIPT 212

Qy    229 LTRFAEALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNVNVESNFITGKGILAI 288
          |:||||||||||||||:|||||||:||::||:||| :||::::|::| :|:||:|||:||
Db    213 LKAYAEALKENSTVKKFSIVGTRSNDPVAFALAEMLKVNKVLKTLNVESNFISGAGILRL 272

Qy    289 MRALQHNTVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSIL 348
          ||||:|||::||||||||||:||:||:|||::|:||:|||::|||||||:|||:||||||
Db    273 VEALPHNTSLVELKIDNQSQPLGNKVEMEIVNMLEKNTTLLKFGYHTQQGPRLRASNAM 332

Qy    349 TRNMDKQRQKRLQE 362
          |||:||:|:|:|
Db    333 MSNNDLVRKRRLAD 346
```

```
RESULT 3
TMOD_HUMAN
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: June 13, 2002, 09:43:59 ; Search time 21.02 Seconds
(without alignments)
2523.374 Million cell updates/sec

Title: US-09-445-362B-4
Perfect score: 2839
Sequence: 1 MSTFGYRRGLSKYESIDEDE.........SIKQLKRVEVPEALRWEHDL 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Database :   PIR_71:*
        1: pir1:*
        2: pir2:*
        3: pir3:*
        4: pir4:*

Searched: 283138 seqs, 96089334 residues

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Total number of hits satisfying chosen parameters: 283138

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	914	32.2	572	2	S18732	autoantigen, 64K -
2	650.5	22.9	359	1	I53091	tropomodulin, - mou
3	644.5	22.7	359	1	A55463	tropomodulin, skel
4	631.5	22.2	359	1	A42336	tropomodulin - hum
5	310.5	10.9	324	2	T25521	hypothetical prote
6	240.5	8.5	1013	2	T46422	hypothetical prote
7	235	8.3	687	2	T29220	hypothetical prote
8	212	7.5	89	2	T46384	hypothetical prote
9	211	7.4	1110	2	I51116	NF-180 - sea lampr
10	203.5	7.2	1621	2	T15264	hypothetical prote
11	203	7.2	1461	2	T41643	probable involveme
12	202.5	7.1	520	2	E97813	WASP, N-WASP, MENA
13	199	7.0	1953	2	S63244	BNI1 protein - yea
14	194	6.8	1657	2	T19536	hypothetical prote
15	193.5	6.8	1375	2	S48375	hypothetical prote
16	192.5	6.8	969	2	T15446	hypothetical prote
17	191	6.7	742	2	F84643	hypothetical prote
18	190.5	6.7	1100	2	T30967	transcription acti
19	189.5	6.7	980	2	S54986	regulatory protein
20	185.5	6.5	474	2	T10271	capsid-associated
21	182.5	6.4	1110	2	T19673	hypothetical prote
22	182	6.4	485	2	A33647	sulfated surface g
23	182	6.4	485	2	A35829	NAB3 protein - yea
24	182	6.4	1634	2	T26517	hypothetical prote
25	181	6.4	2526	2	T20531	hypothetical prote
26	181	6.4	2722	2	T20532	hypothetical prote
27	181	6.4	2738	2	E88320	protein F07A11.6 [
28	180.5	6.4	2562	2	T14266	Xin protein - chic
29	180	6.3	543	2	S25128	61K protein - Auto
30	180	6.3	555	2	T30349	structural protein
31	179.5	6.3	473	2	B85187	glycoprotein homol
32	178.5	6.3	415	1	A34170	acrosin (EC 3.4.21
33	178.5	6.3	505	2	B64560	poly E-rich protei
34	178	6.3	384	2	S51796	vasodilator-stimul
35	176.5	6.2	431	2	S47538	acrosin (EC 3.4.21
36	176.5	6.2	1388	2	T00063	hypothetical prote
37	175.5	6.2	681	2	F85062	hypothetical prote
38	175.5	6.2	599	2	T10798	pherophorin-S - Vo
39	174.5	6.1	421	1	S11674	acrosin (EC 3.4.21
40	174	6.1	1392	2	T51947	probable transcrip
41	173	6.1	907	2	E96636	hypothetical prote
42	172	6.1	440	2	I49681	glyceraldehyde-3-p
43	172	6.1	1206	2	S24407	formin isoform IV
44	172	6.1	5170	2	T15348	hypothetical prote
45	171.5	6.0	1230	2	T18256	probable serine/th

ALIGNMENTS

RESULT 1
S18732
autoantigen, 64K - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 17-Mar-1999
C;Accession: S18732; S26815
R;Dong, Q.; Ludgate, M.; Vassart, G.
J;Clin. Endocrinol. Metab. 72, 1375-1381, 1991
A;Title: Cloning and sequencing of a novel 64-kDa autoantigen recognized by patients
A;Reference number: S18732; MUID:91225220
A;Accession: S18732
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-572 <DON1>
A;Cross-references: EMBL:X54162
R;Dong, Q.H.
submitted to the EMBL Data Library, July 1990
A;Reference number: S26815
A;Molecule type: mRNA
A;Accession: S26815
A;Cross-references: EMBL:X54162 <DON2>
A;Residues: 1-112,'A',114-572 <DON2>
A;Cross-references: EMBL:X54162; NID:g28968; PID:g28969
C;Keywords: glycoprotein
F;25,310,358,386,416/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   32.2%; Score 914; DB 2; Length 572;
Best Local Similarity   34.2%; Pred. No. 3.2e-35;
Matches 220; Conservative 91; Mismatches 135; Indels 198; Gaps 11;

```
Qy     30 LKELERELEDIEPDRNLPVGLRQKSLTEKTPTGTFSREALMAYWEKESQKLLEKE------ 84
          ::||||:||:.:.|| :||||||||:::|||||||:::||||:|:||||:||:::|
Db      1 MEELEKELDVVDPDGSVPVGLRQRNQTEKQSTGVYNREAMLNFCEKETKKLMQREMSMDE 60

Qy     85 -------RLGECGK----- 91
               |: |
Db     61 SKQVETKIDAKNGQERGRDASKKALGPRNSDLGKEPKRGGLKKSFSRDRDESGGKSGEK 120

Qy     92 --------------VAEDKEES------EEELIFTESNSEVS------------- 113
                          :|:| |         |:|||:|:|::
Db    121 PKEEKIIRGIDKGRVRAAVDKKEAGKDGRGGEERAVATKKEEEKKGGDRNTGLSRDKDKKR 180

Qy    114 ---EEVYTEEEEESQEEEEEEDSDEEERTIETAKG------INGTVNYDSVN-- 157
             :|| :|||||:||||:|||:|:||:|:|:|          :||:||:|:
Db    181 EEMKEVAAKKEDDEKVKGERRNTDTRKEGEKMKRAGGNTDMKKEDEKVKRGTGNTDKKDD 240

Qy    158 ------SDNSKPKIFKSQIENINLTNGSNGRNTESPAAIHPCGNPTVIEDAL 203
                : ||| : |:: :::|:|::| ::| || :  | || | | || |
Db    241 EKVKKNEPLHEKEAKDDSKTKTPEKQ-----TPSGPPKPSEGPAKVEEEAAPSIFDEPL 294

Qy    204 DKIRSNDPDTTEVNLNNIENITITQTLTRFAEALKDNTVVKTFSLANTHADDSAAMATAEM 263
          :::|:|||| |||||| || ||||| |||  ||| | |: ||||||| |||||||||||||
```

```
Db  295  ERYKNNDPEMTEVNVNNSDCIINEILVRFTEALEFNTVVKLFPALANTRADDHVAPAIAIM  354
               :  :|| :|| ||:|| ||| : | |  :| |:  ||| :| |
Qy  264  LKANEHITNVNVESNFITGKGILAIMRALQHNTVLTELRFHNQRHIMGSQVEMEIVKLLK   323
         |||| ||| ||||| |||||||||| ||| ||||||||||||| | :||||:|||| |
Db  355  LKANKTITSLNLDSNHITGKGILAIFRALLQNNTLTELRFHNQRHICGGKTEMEIAKLLK   414

Qy  324  ENTTLLRLGYHFELPGPRMSMTSILTRNMDKQRQKRLQEQKQOEGYDGGPNLRTKVWQRG   383
         ||| || |||||| ||||| |::|| |||||||||||||:: :
Db  415  ENTSLLKLGYHFELAGPRMTVINLLSRNMDKQRQKRLQEQRQAQEARKGEKKDLLEVPKAG  474

Qy  384  TPSSSPYVSPRHSPWSSPKLPPKVQTVRSRPLSPVATLPPPPPPPPPPSSQRLPPPPP     443
         |  :| | |  :  |||  |  |                      || :   |||||
Db  475  AVAKG---SPKPSPQPSPK----------------------------PSPKNSPKKGGAPAAPP  507

Qy  444  PPPPLPEKKLITRNIAEVIKQQESAQRALQNGQKKKGKVKKQPNSILKEIKNSLRSV     503
         ||||                                        :|||| |||: ||
Db  508  PPPPL-----------------------------------APPLIMENLKNSLSPA      529

Qy  504  QEKKMEDSSRPSTPQRSAHENLMEAIRGSSSIKOLKRVEVPEALR               547
         ::||| ||          |:  |   | ||| |:||||||| |
Db  530  TQRKMGDKVLPA-QEKNSRDQLLAAIRSSNLKQLKKVEVPKLLQ                 572
```

RESULT 2
I53091

tropomodulin - mouse
C;Species: Mus sp. (mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999
C;Accession: I53091
R;Ito, M.; Swanson, B.; Sussman, M.A.; Kedes, L.; Lyons, G.
Dev. Biol. 167, 317-328, 1995
A;Title: Cloning of tropomodulin cDNA and localization of gene transcripts during mouse
A;Reference number; I53091; MUID:95154574
A;Accession: I53091
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-359 <RES>
A;Cross-references: GB;S76831; NID:g914213; PIDN:AAB33388.1; PID:g914214
C;Genetics:
A;Gene: Tmod
A;Function:
A;Description: modulates the binding of certain isoforms of tropomyosin to actin filamen
C;Superfamily: tropomodulin
C;Keywords: cytoskeleton

```
Query Match             22.9%;  Score  650.5;  DB 1;  Length 359;
Best Local Similarity   40.1%;  Pred. No. 2.6e-23;
Matches  150;  Conservative  66;  Mismatches 111;  Indels  47;  Gaps  8;

Qy    6  YRGLSKYESIDDELLASLSAEELKELERELEDIEPDRN-LPVGLRQKSLTEKPTGTE      64
         || | :||| ||||:|||||| | | |:||||:|  | |||||||||   | || ||
Db    3  YRRELEKYRDLEDEILGALTEEELRTLENELEDLPDNALLPAGLRQKDQTTKAPTGPF      62

Qy   65  SREALMAYWEKESOKLLEKERL----GE-CGKVAEDKEESEEELIFTESNSEVSEEVYTE   119
         ||| :  :|| ||| |||:      ||  ||||:||||:||||| |  |: | : ||
Db   63  KREELLDHLEKQAKEFKDREDLVPYTGEKRGKVWVPKQKPMDPVL------ESVTLE      113

Qy  120  EEEEESQEEEEEEDSDE-----EERIETAKGINGTVNDSVSNDNSKPKIFKS           168
         ||||| |||||  :: |       ||:||:|| ||| :  :|: |||||:
Db  114  PELEEALANASDAELCDIAAILGMHTLMSNQQYQALGSSSIVNKEGLNS-VIKPQYKP     172

Qy  169  QIENLNLTNGSNGRNTESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQT   228
            :| |  |:  :|    |  |      |:|:  ::|   || | :|:| |: |:|
Db  173  VPD---------EEP-----NSTDVEETLERIKNNDPELEEVNLNRINIPIPT           212

Qy  229  LTRPAEAALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNVNVESNFITGKGILAI  288
          || |||||::|:||| |::  |:|: :|   ||||| :|:  |:| ||||| ||:|: |
Db  213  LKAVAEALKENSVVKKFSIVGTRSNDPVAFALAEMLKVNKVLKTLNVESNFISGAGLRL     272

Qy  289  MRALQHNTVLTELRFHNQRHIMGSSQVEMEIVKLLRLGYHELPGPRMSMTSIL         348
          .||:|||::|||:: :|   | |  |||||| :|| :||| | |||| || |
Db  273  VEALPHNTSLVELKIDNQSOPLGNKVEMEIVNMLEKNTTLLKFGYHFTQQGPRLRASNAM   332

Qy  349  TRNMDKQRQKRLQE                                               362
          :|  |  |:|| |
Db  333  MSNNDLVRKRRLAD                                               346
```

RESULT 3
A55463

tropomodulin, skeletal muscle - chicken
C;Species: Gallus gallus (chicken)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 22-Jun-1999
C;Accession: A55463
R;Babcock, G.G.; Fowler, V.M.
J. Biol. Chem. 269, 27510-27518, 1994
A;Title: Isoform-specific interaction of tropomodulin with skeletal muscle and erythr
A;Reference number: A55463; MUID:95050495
A;Accession: A55463
A;Status: Preliminary
A;Molecule type: mRNA
A;Residues: 1-359 <RES>
A;Cross-references: GB;L36678; NID:g562262; PIDN:AAC14459.1; PID:g562263
A;Function:
A;Description: modulates the binding of certain isoforms of tropomyosin to actin fila
C;Superfamily: tropomodulin
C;Keywords: cytoskeleton; skeletal muscle

```
Query Match             22.7%;  Score  644.5;  DB 1;  Length 359;
Best Local Similarity   40.4%;  Pred. No. 4.8e-23;
Matches  150;  Conservative  63;  Mismatches 117;  Indels  41;  Gaps  8;

Qy    6  YRGLSKYESIDDELLASLSAEELKELERELEDIEPDRN-LPVGLRQKSLTEKPTGTE      64
         ||| :|| | |||:|| || ||| | |:|||| | |   | ||||||||   | |||||
Db    3  YRKELEKYRDLEDEKILGALTEEELRKLENELEELDPDNALLPAGLRQKDQTQKPPTGPF    62

Qy   65  SREALMAYWEKESOKLLEKERL----GE-CGKVAEDKEESEEELIFTESNSEVSEEVYTE   119
         ||| |  :|| || | ||:       ||  ||||:|||:: ||:| |  |: | : ||
Db   63  KREELMAHLEQQAKDIKDREDLVPFTGEKRGKAWIPKQKPMDPVL------ESVTLE      113

Qy  120  EEEEESQEEEEEEDSDEEEERTIETAKGINGTVN-----YDSVNSDNSKPKIFKSQIENINL 175
         ||||| |||||  :   :  ||:||:|| ||||        :|: |||||:
Db  114  PELEEALANA-----SDAELCDIAAILGMHTLMSNQQYYEALGSS----------TI      155

Qy  176  TNGSNGRNTESPAAIHPC---GNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTR    231
         .|  : |::  |  |  |       |:|:  ::|::| || | :|:| |:|::| | |
Db  156  VNKEGLNSVIKPTKYKPVPDEEPNSTDVEETLKRIQNNDPDLEEVNLNMINIPVPTLKA    215

Qy  232  FAEALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNVNVESNFITGKGILAIMRA   291
          |||||:|:||| |::  |:|: :|   ||||| :|:  | | ||||| ||:|  |: |
Db  216  LAEALKTNTVVKFSIVGTRSNDPVAFALAEMLKVNNTLKSLNVESNFISGSGLALVEA     275

Qy  292  LQHNTVLTELRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSILTRN   351
         ||.||::|||:: :|   | | ||||||::|:|||||| ||||| | |||| || | |:
Db  276  LQSNTSLIELRIDNQSOPLGNNVEMEIANMLEKNTTLLKFGYHFTQQGPRLRASNAMMNN   335

Qy  352  MDKQRQKRLQE                                                  362
          :|  |:|| |
Db  336  NDLVRKRRLAE                                                  346
```

RESULT 4
A42336

tropomodulin - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A42336
R;Sung, L.A.; Fowler, V.M.; Lambert, K.; Sussman, M.A.; Karr, D.; Chien, S.
J. Biol. Chem. 267, 2616-2621, 1992
A;Title: Molecular cloning and characterization of human fetal liver tropomodulin. A
A;Reference number: A42336; MUID:92129352
A;Accession: A42336
A;Molecule type: mRNA
A;Residues: 1-359 <SUN>

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: June 13, 2002, 09:43:39 ; Search time 14.61 Seconds
922.856 Million cell updates/sec
(without alignments)

Title: US-09-445-362B-4
Perfect score: 2839
Sequence: 1 MSTFGYRGLSKYESIDEDE..........SIKQLKRVEVPEALRWEHDL 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	197	6.9	1105	4	US-08-999-774A-2	Sequence 2, Appl
2	182.5	6.4	1315	3	US-08-899-595-3	Sequence 3, Appl
3	180.5	6.4	1248	4	US-09-080-897-2	Sequence 2, Appl
4	180.5	6.4	1248	4	US-09-323-735-2	Sequence 2, Appl
5	171	6.0	380	2	US-09-026-587-4	Sequence 4, Appl
6	171	6.0	380	2	US-09-227-420-4	Sequence 2, Appl
7	164	5.8	1780	1	US-08-769-309A-5	Sequence 5, Appl
8	164	5.8	1780	3	US-08-994-570-5	Sequence 5, Appl
9	158.5	5.6	418	2	US-09-026-587-1	Sequence 1, Appl
10	158.5	5.6	418	2	US-09-227-420-1	Sequence 1, Appl
11	158	5.6	418	5	US-08-227-536-2	Sequence 2, Appl
12	158	5.6	2414	4	PCT-US95-04682-2	Sequence 2, Appl
13	156.5	5.5	513	4	US-09-041-886-28	Sequence 28, Appl
14	156.5	5.5	530	4	US-09-041-886-29	Sequence 29, Appl
15	156.5	5.5	552	4	US-09-041-886-30	Sequence 30, Appl
16	156.5	5.5	589	4	US-09-041-886-31	Sequence 31, Appl
17	156.5	5.5	3144	2	US-08-246-982A-6	Sequence 6, Appl
18	156.5	5.5	3144	2	US-08-453-265-6	Sequence 6, Appl
19	156.5	5.5	3144	1	US-08-556-419-21	Sequence 21, Appl
20	156.5	5.5	3144	2	US-08-457-273B-42	Sequence 42, Appl
21	156.5	5.5	3144	4	US-09-041-886-15	Sequence 15, Appl
22	155.5	5.5	1255	2	US-09-080-897-4	Sequence 4, Appli
23	155.5	5.5	1255	1	US-08-899-595-1	Sequence 1, Appl
24	155.5	5.5	1255	4	US-09-323-735-4	Sequence 4, Appli
25	152.5	5.4	393	2	US-09-026-587-3	Sequence 3, Appli
26	152.5	5.4	393	2	US-09-227-420-3	Sequence 3, Appli
27	151	5.3	3119	1	US-08-246-982A-16	Sequence 16, Appl
28	151	5.3	3119	1	US-08-453-265-16	Sequence 16, Appl
29	149.5	5.3	905	2	US-08-574-959A-9	Sequence 9, Appl
30	149.5	5.3	905	4	US-09-357-014-9	Sequence 9, Appl
31	149.5	5.3	1135	2	US-08-574-959A-7	Sequence 7, Appl
32	149.5	5.3	1135	4	US-09-357-014-7	Sequence 7, Appl
33	149	5.2	323	4	US-09-029-213B-25	Sequence 25, Appl
34	149	5.2	1805	1	US-07-853-913-2	Sequence 2, Appl
35	149	5.2	1958	1	US-07-945-283-2	Sequence 2, Appl
36	149	5.2	2441	1	US-08-194-468-2	Sequence 2, Appl
37	149	5.2	2441	1	US-08-961-739-2	Sequence 2, Appl
38	147	5.2	281	2	US-08-810-453-2	Sequence 2, Appli
39	147	5.2	281	3	US-08-815-190A-2	Sequence 2, Appli
40	147	5.2	281	4	US-09-290-640-25	Sequence 25, Appli
41	147	5.2	281	4	US-09-479-524-3	Sequence 3, Appli
42	147	5.2	281	4	US-08-339-214-8	Sequence 8, Appli
43	147	5.2	281	5	US-08-339-214-30	Sequence 30, Appl
44	147	5.2	281	5	PCT-US95-00362-2	Sequence 2, Appli
45	147	5.2	976	4	US-09-104-324B-4	Sequence 4, Appli

RESULT 1
US-08-999-774A-2
Sequence 2, Application US/08999774A

; Patent No. 6274312
GENERAL INFORMATION:
APPLICANT: Gish, Kurt C.
APPLICANT: Seghezzi, Wolfgang
APPLICANT: Shanahan, Frances
APPLICANT: Lees, Emma M.
APPLICANT: McClanahan, Terrill K.
TITLE OF INVENTION: Intracellular Regulatory Molecules;
TITLE OF INVENTION: Related Reagents
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/999,774A
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/032,818
FILING DATE: 11-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0646
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1105 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-999-774A-2

Query Match 6.9%; Score 197; DB 4; Length 1105;
Best Local Similarity 20.0%; Pred. No. 1e-05;

ALIGNMENTS

RESULT 2
US-08-899-595-3
Sequence 3, Application US/0889595
Patent No. 6111072
GENERAL INFORMATION:
 APPLICANT: Takahashi, Shuh
 APPLICANT: Narumiya, No. 6111072uaki
 APPLICATION NUMBER: US/08/899,595
 FILING DATE: 24-JUL-1997
 CLASSIFICATION: 435
 CURRENT APPLICATION DATA:
 SOFTWARE: Patentin Release #1.0, Version #1.30
 OPERATING SYSTEM: PC-DOS/MS-DOS
 COMPUTER: IBM PC compatible
 MEDIUM TYPE: Floppy disk
 COMPUTER READABLE FORM:
 GENERAL INFORMATION:
 Patent No. 6111072
 Sequence 3, Application US/0889595

TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
 ENCODING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 2007-5109

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/899,595
PRIOR APPLICATION DATA:
FILING DATE: 26-AUG-1996
APPLICATION NUMBER: JP 8-242701
FILING DATE: 25-MAR-1997
APPLICATION NUMBER: JP 9-90170
ATTORNEY/AGENT INFORMATION:
NAME: Stephen A. Bent

Query Match   99;   Conservative   70;   Mismatches   162;   Indels   164;   Gaps   20;

```
Qy       7  RRGLSKYESIDEDELLASLSA--EELKELERELEDIEPDRNLPVGLRQKSLTEKTPTGTF   64
            |:|  :||  ||:|| :     ||:||||||||||||||||||||||:||||||||||||||
Db     716  KAALEEFSRVREEVPLELVEAHVKKVQEAARASGKVDP----TYGLESSCI--AGTGPD  768

Qy      65  SREALMAYWEKESQKLLEKERLGECGKVAEDKEESEEELIFTESNSEVSEEVYTEEEE  124
            ||:|||||||||||||||||||||| | ||||:|||||:||||||||||||||||||
Db     769  EPEKL----EGAEEKMEADPDQQOPEKAENKVENETD-----EGDKAQDGENEKNSEK  818

Qy     125  SQEEEEEEDSDEEERTIETAKGINGTVNYDSVNSDNSKPKIFKSQIENINLTNGSNGRNT  184
            :|:| |   |:    |  |:||  |   |  | |  ||| | |:|| |:|| ||:|
Db     819  EQDSEVSEDTKSEEKETE-----------------------ENKELT-------  842

Qy     185  ESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTRFAEALKDNTVVKT  244
            :::   |  |  |:  | ||:|| :         ||| ||||:|||  ||||:
Db     843  -------------DICKERESDTGKKKVEH-----------------------  859

Qy     245  FSLANTHADDSAAMAIAEMLKANEHITNVNVESNFITGKGILAIMRALQHNTVLTELR-F  303
            ::|  |  | :  :||||| || | |||  |||: :||  | |:||:||| |:|: |:|
Db     860  -EISEGNVATAAAALASAATKAKHL--AAVEERKI--KSLVALIVETQMKKLEIKLRHF  914

Qy     304  HNQRHIMGSQ----VEMEIVKLLKENTLLRLGYHFELPGPRMSMTSILTRNMDKQRQKRL  360
            |||||||||     |||||||||||||||:|  ||||||||||||: ||||||||||||
Db     915  EELETIMDREKALEQORQQLLTE-----RQNFHME----QLKYAEPLRARQOMEDOQHG-  964

Qy     361  QEQKQOEGYDGGPNLRTKVWQRGTPSSSPIVSPRHSPWSSPPKLP----------------  404
            |||||| |  |||||| |||||||||||||||||||||||||||
Db     965  QNPQOAHQHSGGPCL-APLGAAGHPGMHPHQOPPYPLMHIHQMPPHPPQOPGQIPGPGSM  1023

Qy     405  --------KKVQTVRSRPLSPVATLPPPPPPPPP-------------PPSSQRLP  439
                    ||||||||||||||||||||||||||           |||:||||
Db    1024  MPGQHMPGRMIPIVAAN-IHPSGSGPTPPGMPPMPCNILGPRVPLTAPNGMYPPPPPQOP 1082

Qy     440  PPPPP----PPPLP  450
            ||||    ||||
Db    1083  PPPPPADGVPPPAP 1097
```

REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 049441/0112
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1315 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-595-3

Best Local Similarity   22.0%; Pred. No. 0.00013;
Matches 116; Conservative 70; Mismatches 174;

Query Match   6.48;   Score 182.5; DB 3;   Length 1315;
Indels 167;   Gaps 26;

```
Qy      16  IDEDELLASL-----SAEELKE------LER-ELEDIEPDRNLPVGLRQKSLTEKTPTGTF  64
            |:||||||||      |||||       |:| ||||||||||||||||||||||||||||
Db     300  IDAAKLLSALCILPQPEDMNERVLEAMTERAEMDEVERFQPLLDGIK-SGTTIALKVGCL  358

Qy      65  SR-EALMAYWEKESQKLLEKERLGECG--KVAEDKEESEEELIFTESNSEVSEEVYTEEE  121
            |  ||||||||||||||||||||||    |||||:|||||||||||||||||||||||||
Db     359  QLINALITPAEELDFRVHIRSELMRLGLHQVLQDLREIENE-----DMRVQLNVFDEQG  412

Qy     122  EERSQE------EEEEEDSDEE-ERTIEFAKGINGTVNYDSV-------NSDNSKPK  164
            ||||         ||||||||| |  |:|||  ||| |  |         |||||||
Db     413  EEDSYDLKGRLDDIRMEMDDFENEVFQILLNTVKDSKAEPHFISILQHLLJVRNDYEARPQ 472

Qy     165  IFKSQIENINLTNGSNGRNTESPAAIHPCGNPTVIEDALDKI-----KSNDPDTTEVNLN-  219
            |||||||||||||||||||||:||||:||||||||||||||     |:|||||||||||
Db     473  YYK--------LIEECISQIVLHKNGADPFKCRHLQI  502

Qy     220  NINITTQTLTRFAEALKDNTVVKTFSLANTHADDSAAMAIAEMLKANEHITNVNVESNF  279
            : | |:| |||| |  |||  | : |   |   |   :|||||||||||  |  | |:|
Db     503  EIEGLIDQMIDK------------TKVEKSEAKA-AELEK--------  535

Qy     280  ITGKGILAIMRALQHN---TVLTELR-----FHNQRHIMGSQ--VEMEIVKLLKENTTLL  329
            ||:| | |:||:||||   |:|:|:      |||||||||||   ||||||||||||||
Db     536  TARHELQVEMKKNESDFEQKLQDLQGERKDALHSEKQQIATEKQDLEAEVSQLTGEVAKLT  595

Qy     330  RLGYHFELPGPRMSMTSI----LTRNMDKQRQKRLQEEQKQOEGYDG-------GPNLRTK  378
            ||||||||||||||||:        |||||||||||||:  |   |        ||:|||
Db     596  K-----ELEDAKKEMASLSAAAITVPPSVPSRAPVPPAPPLPGDSCTIIPPPPAPGDST-  649

Qy     379  VWQRGTPSSSPIVSPRHSPW-------SSPKLPKKVQTVRSRPLSPVATLPPPPP----  426
            ||||||||||||||||||          |||||||||||||||||||||||||
Db     650  ----TPPPPPPPPPPLPGGGVCISSPPSLPGGTVAISPPPLSGDATIPPPPPLPEGV  704

Qy     427  --------------PPPPPPSSQRLPPPP--------PPPPPLP  450
                          |||||||||||||         |||||||
Db     705  GIPSPSSLPGGTAIPPPPPPLPGSARIPPPPPPPLPGSAGIPPPPPLP  751
```

RESULT 3
US-09-080-897-2
Sequence 2, Application US/09080897
Patent No. 5985574
GENERAL INFORMATION:
 APPLICANT: King, Mary-Claire
 APPLICANT: Lynch, Eric D.
 APPLICANT: Lee, Ming
 APPLICANT: Morrow, Jan E.
 APPLICANT: Welcsh, Piri L.
 APPLICANT: Leon, Pedro E.
TITLE OF INVENTION: Modulators of Actin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH

OM protein - protein search using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: June 13, 2002, 09:41:49 ; Search time 34.22 Seconds
(without alignments)
1791.722 Million cell updates/sec

Title: US-09-445-362B-4
Perfect score: 2839
Sequence: 1 MSTFGYRRGLSKYESIDEDE.........SIKQLKRVEPEALRWEHDL 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Post-processing: Minimum Match 0%
Maximum Match 100%

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   A.Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Listing first 45 summaries

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2839	100.0	552	20	AAW90172	Human heart muscle
2	2839	100.0	552	22	AAB86278	Human DCMAG-1 prot
3	2063.5	72.7	531	22	AAU19573	Human diagnostic a
4	1447	51.0	302	21	AAB42655	Human ORFX ORF2419
5	964	34.0	600	22	AAM38842	Human polypeptide
6	963	33.9	610	22	AAM40628	Human polypeptide
7	779.5	27.5	358	21	AAY95848	Autoantigen diagno
8	709.5	25.0	330	21	AAY95846	Autoantigen diagno
9	705	24.8	168	21	AAB43003	Human ORFX ORF338
10	650.5	22.9	351	22	AAM79201	Human protein SEQ
11	613	21.6	352	22	AAM39533	Human polypeptide

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

12	613	21.6	352	22	AAB98824	Human tropomodulin
13	613	21.6	358	22	AAM41319	Human polypeptide
14	561	19.8	368	21	AAB43100	Human ORFX ORF2864
15	553.5	19.5	391	22	AAU30836	Novel human secret
16	480	16.9	104	22	AAU21876	Human cardiovascul
17	416	14.7	223	22	AAM80185	Human protein SEQ
18	378	13.3	92	20	AAW90173	Human heart muscle
19	376	13.2	174	22	AAM25794	Human protein sequ
20	374	13.2	367	22	AAB66832	Drosophila melanog
21	374	13.2	403	22	AAB58218	Drosophila melanog
22	360	12.7	273	22	AAO11834	Human polypeptide
23	348	12.3	141	22	AAU30835	Novel human secret
24	240.5	8.5	1017	22	AAM40352	Human polypeptide
25	226	8.0	1174	22	AAG85039	Shrimp white spot
26	211.5	7.4	560	22	AAB95239	Human protein sequ
27	198.5	7.0	1162	22	AAB62516	Drosophila melanog
28	197	6.9	1105	22	AAB09325	Drosophila melanog
29	195.5	6.9	527	22	AAB58220	Human intracellula
30	194.5	6.9	983	20	AAY09513	Drosophila melanog
31	194.5	6.9	3111	22	AAB60327	Mouse JMY protein
32	193.5	6.8	757	22	AAG85017	Drosophila melanog
33	193	6.8	1700	22	AAB64608	Shrimp white spot
34	187	6.6	635	19	AAW85034	Green fluorescent
35	187	6.6	635	21	AAY70781	Drosophila melanog
36	186	6.6	3536	22	AAB65480	EGFP-VASP fusion p
37	184	6.5	748	22	AAM58340	Human brain expres
38	184	6.5	3263	22	AAB67210	Drosophila melanog
39	183.5	6.5	2819	22	AAB35408	Drosophila melanog
40	182.5	6.4	1315	20	AAW76734	Human 07CG27 gene
41	182	6.4	100	21	AAG00851	Human mDia Rho tar
42	181	6.4	2518	21	AAB40574	Human secreted pro
43	180.5	6.4	473	21	AAG24963	Human ORFX ORF338
44	180.5	6.4	1248	20	AAY13464	Arabidopsis thalia
45	180.5	6.4	2724	22	ABG20119	Novel human diagno

RESULT 1
AAW90172

ID   AAW90172 standard; Protein; 552 AA.
XX
AC   AAW90172;
XX
DT   16-MAR-1999   (first entry)
XX
DE   Human heart muscle specific protein.
XX
KW   Heart muscle; cardiac insufficiency; detection; interactor; modulator;
KW   heart disease; diagnostic; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9856907-A1.
XX
PD   17-DEC-1998.
XX
PF   15-JUN-1998;   98WO-EP03584.
XX
PR   13-JUN-1997;   97DE-1025186.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Domdey H,   Henkel T,   Hofmann ME;
XX
DR   WPI; 1999-060330/05.
DR   N-PSDB; AAV74263, AAV74264, AAV74365.
XX
PT   New nucleic acid selectively expressed in insufficient cardiac
PT   muscle - for treatment and diagnosis of heart disease
XX
PS   Claim 1; Fig 4; 58pp; German.

XX
CC   This sequence represents a fragment of a human heart muscle specific
CC   protein. This sequence is used for treating heart disease, especially
CC   cardiac insufficiency, and for detecting interactors (e.g. in a
CC   two-hybrid assay, identifying polypeptides that interact with the
CC   sequence, potential therapeutic agents, or compounds that modulate
CC   such interactions). They can also be used diagnostically, e.g. in
CC   standard amplification, Northern blotting or immunoassays. This sequence
CC   has been identified in human cardiac tissue cDNA libraries and is more
CC   strongly expressed in diseased tissue than in normal tissues, suggesting
CC   that it is casually related to cardiac insufficiency. The new materials
CC   provide specific and accurate diagnosis and therapy of heart disease.
XX
SQ   Sequence   552 AA;

Query Match   100.0%; Score 2839; DB 20; Length 552;
Best Local Similarity   100.0%; Pred. No. 1.9e-169;
Matches 552; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MSTFGYRRGLSKYESIDEDELLASLSAEELKELERELEDIEPDRNLPVGLRQKSLTEKTP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mstfgyrrglskyesidedellaslsaeelkelerelediepdrnlpvglrqksltektp 60

Qy   61 TGTFSREALMAYWEKESQKLLEKERLGECGKVAEDKDEESEEELIFTESNSEVSEEVYTEE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 tgtfsrealmaywekesqkllekerlgecgkvaedkdeeseeeliftesnsevseevytee 120

Qy   121 ·EEEESQEEEEEDSDEEERTIETAKGINGTVNYDSVNSDNSKPKIFKSQIENINLTNGSN 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 eeeesqeeeeedsdeeertietakgingtvnydsvnsdnskpkifksqieninltngsn 180

Qy   181 GRNTESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTRFAEALKDNT 240
          ·||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 grntespaaihpcgnptviedaldkiksndpdttevnlnnienittqtltrfaeealkdnt 240

Qy   241 VVKTFSLANTHADDSAAAMAIAEMLKANEHITNVNVESNFITGKGILAIMRALQHNTVLTE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 vvktfslanthaddsaaamaiaemlkanehitnvnvesnfitgkgilaimralqhntvlte 300

Qy   301 LRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSILTRNMDKQRQRRL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 lrfhnqrhimgsqvemeivkllkenttllrlgyhfelpgprmsmtsiltrnmdkqrqkrl 360

Qy   361 QEOKQQEGYDGGPNLRTKVWQRGTPSSSSPYVSPRHSPWSSSPKLPEKKVQTVRSRPLSPVAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 qeqkqqegydggpnlrtkvwqrgtpsssspyvsprhspwssspklpkkvqtvrsrplspvat 420

Qy   421 LPPPPPPPPPPPPPPPSSQRLPPPPPPPPPLPEKKLITRNIAEVIKQQESAQRALQNGQKKK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 lpppppppppppppppssqrlpppppppplpekklitrniaeviikqqesaqralqngqkkk 480

Qy   481 KGKKVKKQPNSILKEIKNSLRSVQEKKMEDSSRPSTPQRSAHENLMEAIRGSSIKQLKRV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 kgkkvkkqpnsilkeiknslrsvqekkmedssrpstpqrsahenlmeairgssikqlkrv 540

Qy   541 EVPEALRWEHDL 552
          ||||||||||||
Db   541 evpealrwehdl 552

RESULT   2

AAB86278

ID   AAB86278 standard; protein; 552 AA.
XX
AC   AAB86278;
XX
DT   06-SEP-2001 (first entry)
XX
DE   Human DCMAG-1 protein.
XX
KW   DCMAG-1; human; dilative cardiomyopathy; cardiac muscle cell;
KW   cardioactive agent; heart disease; treatment.

XX
OS   Homo sapiens.
XX
PN   WO200146388-A2.
XX
PD   28-JUN-2001.
XX
PF   21-DEC-2000; 2000WO-EP13101.
XX
PR   22-DEC-1999; 99DE-1062154.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Roenicke V, Nave B, Henkel T;
XX
DR   WPI; 2001-418054/44.
XX
PT   New pathologically altered cardiomyocyte, useful for identifying and
PT   detecting cardioactive agents, produced by stimulating healthy cells
PT   with hormones and cytokines -
XX
PS   Claim 15; Page 38-40; 40pp; German.
XX
CC   This invention describes a novel pathologically altered cardiac muscle
CC   cell (A), isolated from healthy cardiac tissue and/or cells by
CC   stimulation of isolated cells with suitable hormones (or analogs) and/or
CC   cytokines. (A) are used to identify or detect cardioactive agents,
CC   potentially useful for prevention or treatment of heart disease, also to
CC   study molecular changes that cause heart disease. This sequence
CC   represents the human DCMAG-1 protein, associated with dilattive
CC   cardiomyopathy which is used to illustrate the method of the invention.
XX
SQ   Sequence   552 AA;

Query Match   100.0%; Score 2839; DB 22; Length 552;
Best Local Similarity   100.0%; Pred. No. 1.9e-169;
Matches 552; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MSTFGYRRGLSKYESIDEDELLASLSAEELKELERELEDIEPDRNLPVGLRQKSLTEKTP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mstfgyrrglskyesidedellaslsaeelkelerelediepdrnlpvglrqksltektp 60

Qy   61 TGTFSREALMAYWEKESQKLLEKERLGECGKVAEDKDEESEEELIFTESNSEVSEEVYTEE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 tgtfsrealmaywekesqkllekerlgecgkvaedkdeeseeeliftesnsevseevytee 120

Qy   121 EEEESQEEEEEDSDEEERTIETAKGINGTVNYDSVNSDNSKPKIFKSQIENINLTNGSN 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 eeeesqeeeeedsdeeertietakgingtvnydsvnsdnskpkifksqieninltngsn 180

Qy   181 GRNTESPAAIHPCGNPTVIEDALDKIKSNDPDTTEVNLNNIENITTQTLTRFAEALKDNT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 grntespaaihpcgnptviedaldkiksndpdttevnlnnienittqtltrfaeealkdnt 240

Qy   241 VVKTFSLANTHADDSAAAMAIAEMLKANEHITNVNVESNFITGKGILAIMRALQHNTVLTE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 vvktfslanthaddsaaamaiaemlkanehitnvnvesnfitgkgilaimralqhntvlte 300

Qy   301 LRFHNQRHIMGSQVEMEIVKLLKENTTLLRLGYHFELPGPRMSMTSILTRNMDKQRQRRL 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 lrfhnqrhimgsqvemeivkllkenttllrlgyhfelpgprmsmtsiltrnmdkqrqkrl 360

Qy   361 QEOKQQEGYDGGPNLRTKVWQRGTPSSSSPYVSPRHSPWSSSPKLPEKKVQTVRSRPLSPVAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 qeqkqqegydggpnlrtkvwqrgtpsssspyvsprhspwssspklpkkvqtvrsrplspvat 420

Qy   421 LPPFPPPPPPPPPPPPSSQRLPPPPPPPPPLPEKKLITRNIAEVIKQQESAQRALQNGQKKK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 lpppppppppppppppssqrlpppppppplpekklitrniaeviikqqesaqralqngqkkk 480

Qy   481 KGKKVKKQPNSILKEIKNSLRSVQEKKMEDSSRPSTPQRSAHENLMEAIRGSSIKQLKRV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||